

12-4-2006

The Association of Hypertension Diagnosis with Smoking Cessation: Application of Multiple Logistic Regression Using Biostatistical and Epidemiological Methods

LaTonia Clay

Follow this and additional works at: https://scholarworks.gsu.edu/math_theses



Part of the [Mathematics Commons](#)

Recommended Citation

Clay, LaTonia, "The Association of Hypertension Diagnosis with Smoking Cessation: Application of Multiple Logistic Regression Using Biostatistical and Epidemiological Methods." Thesis, Georgia State University, 2006.
https://scholarworks.gsu.edu/math_theses/18

This Thesis is brought to you for free and open access by the Department of Mathematics and Statistics at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Mathematics Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

The Association of Hypertension Diagnosis with Smoking Cessation:
Application of Multiple Logistic Regression Using Biostatistical and Epidemiological Methods

by

LaTonia A. Clay

Under the Direction of Yu-Sheng Hsu, PhD.

ABSTRACT

Hypertension and smoking are two major issues threatening the nation's health. Previous studies examining their relationship have resulted in conflicting reports. The aim of this study is to determine if a relationship exists between smoking cessation and hypertension diagnosis. Data from the Third National Health and Nutrition Examination Survey (NHANES III) were used in this investigation. Physical examination measurements of blood pressure and self-reported diagnosis and smoking behavior were used to define hypertension and smoking status. The odds of prior hypertension diagnosis associated with smoking cessation was estimated from a multivariate logistic regression model, adjusting for gender, age, ethnicity, BMI, physical activity, HDL cholesterol, and alcohol use. Results of the investigation found that unsuccessful smoking cessation was associated with a decreased odds of prior hypertension diagnosis, adjusting for the presence of confounders ($OR=0.816$, $p<.001$). Thus, hypertension diagnosis may indeed lead to the decision to quit smoking. Future studies on this finding are encouraged.

INDEX WORDS: hypertension, smoking, cessation, diagnosis, logistic regression, NHANES

THE ASSOCIATION OF HYPERTENSION DIAGNOSIS WITH SMOKING CESSATION:
APPLICATION OF MULTIPLE LOGISTIC REGRESSION USING BIOSTATISTICAL
AND EPIDEMIOLOGICAL METHODS

By

LaTonia A. Clay

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of:

Master of Science

in the College of Arts and Sciences

Georgia State University

2006

Copyright by
LaTonia A. Clay
2006

THE ASSOCIATION OF HYPERTENSION DIAGNOSIS WITH SMOKING CESSATION:
APPLICATION OF MULTIPLE LOGISTIC REGRESSION USING BIOSTATISTICAL
AND EPIDEMIOLOGICAL METHODS

By

LaTonia A. Clay

Major Professor: Yu-Sheng Hsu, PhD.

Committee: Gengsheng (Jeff) Qin, PhD.

Xu Zhang, PhD.

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
December 2006

ACKNOWLEDGEMENTS

I wish to express my gratitude to Dr. Yu-Sheng Hsu for accepting me as a graduate student in the Department of Mathematics and Statistics at Georgia State University, for his guidance and support during my studies, and for teaching me everything I know about conducting statistical analyses. I am especially thankful to my thesis committee members Dr. Gingsheng (Jeff) Qin and Dr. Xu Zhang for teaching me subject matter relevant to this document and for their continued support during my studies. I would also like to thank all other faculty and staff of the Mathematics and Statistics Department of Georgia State University for making my time within the department an enjoyable experience. I also wish to thank Dr. Michael P. Eriksen for accepting me as a graduate student in the Institute of Public Health at Georgia State University and for commenting on earlier drafts of this document. I also thank Dr. Solomon Ike Okoson for his editorial reviews of earlier drafts of this document and, most importantly, for providing me with the data used in the analyses of this study.

TABLE OF CONTENTS

| | |
|--|------|
| Acknowledgements | iv |
| List of Tables | vii |
| List of Figures | viii |
| Chapter | |
| 1. Introduction | 1 |
| Cigarette Smoking as a Public Health Concern | 1 |
| Hypertension as a Public Health Concern | 5 |
| Purpose and Hypotheses | 7 |
| Gaps in Current Knowledge | 9 |
| 2. Review of Literature | 12 |
| Evidence for the link between smoking and hypertension | 12 |
| Evidence for the interaction of smoking and hypertension | 12 |
| Evidence for inequalities in smoking and hypertension | 13 |
| 3. Methods and Procedures | 16 |
| Data Source | 16 |
| Study Population | 17 |
| Outcome definitions | 17 |
| Potential Confounders | 18 |

| | |
|--|----|
| Statistical Analyses: Biostatistical Method | 20 |
| Statistical Analyses: Epidemiological Method | 22 |
| 4. Results | 24 |
| Descriptive Characteristics | 24 |
| Hypothesis Testing | 28 |
| 5. Discussion and Conclusion | 48 |
| 6. References | 52 |
| 7. Appendix: SAS Codes | 59 |

LIST OF TABLES

| | |
|---|----|
| Table 1: Characteristics of the Study Population | 25 |
| Table 2: Mean of Continuous Variables | 26 |
| Table 3: Normality Check for Continuous Variables | 26 |
| Table 4: Collinearity Check for Continuous Variables | 27 |
| Table 5: Testing the Association of the Categorical Variables to the Dependent Variable | 28 |
| Table 6: Testing the Association of the Continuous Variables to the Dependent Variable | 28 |
| Table 7: Univariable Analyses | 29 |
| Table 8: Fitting a Preliminary Main Effects Model | 29 |
| Table 9: Results of the Fractional Polynomial Transformation | 33 |
| Table 10: Interactions Significant at the .05 level | 34 |
| Table 11: Fitting a Model Containing the Main Effects and Significant Interactions | 34 |
| Table 12: Results of Stepwise Regression | 36 |
| Table 13: Epidemiological Approach—Univariable Analyses | 37 |
| Table 14: Epidemiological Approach—Fitting a Preliminary Main Effects Model | 38 |
| Table 15: Fitting a Model Containing Main Effects and Significant Interaction Terms | 39 |
| Table 16: Epidemiological Approach Results of Stepwise Regression | 40 |
| Table 17: Epidemiological Approach—Assessing the Fit of the Model | 41 |
| Table 18: Epidemiological Approach—Final Model Interpretations | 45 |

LIST OF FIGURES

| | |
|--|----|
| Figure 1: Smoothed Scatter plot of Patients Diagnosed vs. Age | 30 |
| Figure 2: Smoothed Scatter plot of Patients Diagnosed vs. Systolic Blood Pressure | 30 |
| Figure 3: Smoothed Scatter plot of Patients Diagnosed vs. BMI | 31 |
| Figure 4: Smoothed Scatter plot of Patients Diagnosed vs. Diastolic Blood Pressure | 31 |
| Figure 5: Smoothed Scatter plot of Patients Diagnosed vs. HDL cholesterol | 32 |
| Figure 6: Epidemiological Approach--Sensitivity vs. Probability Level Plot | 42 |
| Figure 7: Epidemiological Approach--ROC Curve (Sensitivity vs. 1-Specificity Plot) | 43 |
| Figure 8: Epidemiological Approach: Delta Chi-Square vs. Estimated Probability | 43 |
| Figure 9: Epidemiological Approach: Delta Deviance vs. Estimated Probability | 44 |
| Figure 10: Epidemiological Approach: Delta Beta Hat vs. Estimated Probability | 44 |

Chapter I: Introduction

The publication of the first Surgeon General's Report in 1964 raised public concern about the health effects of cigarette smoking (United States, 2004). Subsequently, the harmful effects of smoking have remained a major public health issue in the 21st century. Specifically, cigarette smoking is currently the leading cause of preventable mortalities in the United States, responsible for nearly half a million deaths per year (CDC, 2005). Smoking is also a leading cause of chronic disease morbidities, affecting 8.6 million people per year (Hyland et. al, 2004). Some of the chronic diseases attributed to smoking include: pancreatic cancer, stomach cancer, respiratory diseases such as chronic obstructive pulmonary disease and pneumonia, and cardiovascular diseases such as atherosclerosis, and coronary heart disease (United States, 2004). Of particular interest to this thesis is the risk of cardiovascular disease attributed to smoking. Primarily, this thesis seeks to examine the role of hypertension as a precursor to the development of smoking-induced cardiovascular disease. Moreover, it seeks to provide a better understanding of the link between smoking cessation and hypertension diagnosis.

Cigarette Smoking as a Public Health Concern

Before developing an understanding of the link between smoking cessation and hypertension, it is important to have an understanding of both smoking and hypertension as independent public health issues. In addition to the burden of chronic disease morbidities and mortalities previously discussed, cigarette smoking is also a public health issue due to its addictive chemicals that prevent smokers from quitting. Most tobacco products, including

cigarettes, contain a potent chemical known as nicotine. The presence of nicotine in cigarettes makes cigarette use a dependence disorder (Galvin, 1992). For this reason, many smokers find quitting very difficult. Studies reveal that over 60% of smokers have a desire to quit, yet among those who actually attempt to quit, more than half are often unsuccessful (CDC, 2005). The addictive nature of cigarette smoking not only makes it difficult for smokers to quit, but also makes public health efforts promoting smoking cessation difficult to achieve.

Several factors predict successful quit attempts and cessation among smokers. One factor is the presence of a strong non-smoking role model, especially one residing in the same home as the smoker (Gritz et. al, 2003). A 2005 study of predictors of attempted quitting and cessation among adult smokers concluded that less exposure to other smokers in general during quit attempts is one of the keys to successful quitting (Tucker et. al, 2005). Another factor that the study found to be a predictor of successful quitting is quitting during the transition into parenthood. Smokers with increased feelings of responsibility and obligation to others were more likely to be successful at quitting, whereas attempting after this transition posed a potential barrier to quitting (Gritz et. al, 2003). One other significant predictor of successful quitting was perceived health status. Specifically, the study concluded that smokers were more likely to quit if they were motivated to adopt a healthier lifestyle or if they strongly believed that there were health benefits of quitting (Tucker et. al, 2005).

There are various methods of quitting. Although some smokers have tried to quit with the help of nicotine replacement therapy or other types of pharmacotherapy, most people (69%) have tried to quit on their own (McGrady & Pederson, 2002). In terms of self-help attempts, social, physiological, and psychological factors collectively motivate individuals to quit and resist the temptation of relapse (King et. al, 1999). However, pharmacotherapy has been found

to be more successful in helping smokers quit (Green et. al, 2000). According to national data on smoking cessation, those who are more likely to be successful at quitting are those who are of White race, older age, higher income, and smoke fewer numbers of cigarettes per day (Hyland et. al, 2004). Hence, public health efforts in promoting smoking cessation may be more successful in individuals who do not meet these characteristics.

Recent public health efforts to encourage smoking cessation have included public policies on tobacco control, excise taxes, sales restrictions, community-based interventions, and other multidisciplinary strategies (King et. al, 1999). According to the Guide to Community Preventive Services, many of these strategies provide strong evidence for effectiveness. For example, in seven of eight studies concerning the effectiveness of increasing the price of tobacco products, increasing the price resulted in decreases in both the number of people who use tobacco and the quantity they consume (Guide to Community Preventive Services, 2003). Mass media campaigns have also been effective strategies at reducing consumption and increasing cessation among tobacco users. Provider education has also shown to be effective strategies to increase cessation among patients receiving advice and education in healthcare settings (Guide to Community Preventive Services, 2003). Similarly, telephone counseling and tobacco quit lines have also provided strong evidence for the effectiveness of increasing cessation.

Despite the achievements of various smoking cessation methods, there are several barriers to smoking cessation, both external and internal. A key external barrier is the environment in which cigarette smoking is advertised and, quite often, encouraged. Although often viewed as a catalyst for smoking initiation, cigarette marketing may also serve as an impediment to cessation efforts. Marketing strategies combined with the addictive chemicals found in tobacco products could increase the desire to smoke such that the overall effect on the

smoker is both psychological and physiological in nature (Biener & Siegel, 2000). Other external factors associated with influencing cessation behavior include family, peers, and the presence of another smoker. Specifically, these groups of people may hinder smoking cessation behavior by engaging in the behavior themselves or encouraging participation in the behavior.

Internal barriers to smoking cessation include cognitive factors such as people's beliefs about their smoking behavior. Results of a 1992 study concerning beliefs about smoking and actual smoking behavior revealed that many smokers smoked as a result of what they believed to be rewards of smoking. These "rewards" included anxiety-reduction, stimulation, social reward, pleasurable taste, and something to "keep the hands busy" (Galvin, 1992). These "rewards" justify smoking and reinforce the desire to continue the behavior. In this way, this can be a barrier to smoking cessation.

Another commonly held belief among many smokers, especially women, is that weight gain is a consequence of quitting smoking. This belief is based on scientific evidence indicating that stopping smoking commonly leads to weight gain (Niskanen, Laaksonen et al. 2004). The most obvious mechanism by which this occurs is that ex-smokers often substitute cigarettes for food, leading to increased weight. Another possible explanation is that nicotine may affect the regulation of food intake by the hypothalamus such that smokers feel less inclined to eat in comparison to non-smokers. Thus, smoking cessation reverses this regulatory process and increases food intake (Janzon, Hedblad et al. 2004).

Weight gain is, in fact, a well-established risk factor for hypertension. In this regard, the weight increase experienced by many smokers who quit could, in fact, increase the incidence of hypertension among this group. This would be antithetical to the goals of public health cessation efforts by potentially reducing the positive effects of quitting. As a rebuttal to this argument,

studies have shown that ex-smokers who develop hypertension have a lower risk of cardiovascular disease than those who continue to smoke; among women, the risk is lower in comparison to both hypertensive smokers and smokers with normal blood pressure (Janzon, Hedblad et al. 2004). Thus, the increased risk of hypertension does not offset the positive effects of cessation, which include reduced mortality for all causes of cancer and ischaemic heart disease (Wen et. al, 2005). The interest of this thesis is to further explore this relationship between hypertension and smoking cessation.

Hypertension as a Public Health Concern

Hypertension, often called a “silent killer”, is the leading risk factor for cardiovascular disease, responsible for approximately three million annual deaths worldwide and about 3% of the global health care expenditures (Pardell et. al, 1998). Consequences of hypertension include heart disease and stroke, end-stage renal disease, and peripheral vascular disease (Os et. al, 2004). Approximately one in four adults in the United States has hypertension (Centers for Disease Control and Prevention 2005).

Hypertension is defined as having a blood pressure of 140/90 mmHg or more (Pardell et. al, 1998). The top number (systolic blood pressure) appears to be more predictive of future hypertension status than the bottom number (diastolic blood pressure) (Vasan et. al, 2001). Although the consequences of hypertension can be life-threatening, it is possible for patients to control their hypertension. However, control of the condition requires alterations in diet and lifestyle such as decreasing salt intake, increasing the amount of fruits and vegetables in the diet, increasing physical activity, and weight loss (Cronin, 1986).

Until recently, blood pressure was considered “normal” if it was maintained below 140/90 mmHg and “optimal” if it was 120/80 mmHg (Miller & Jehn, 2004). However, new guidelines created by the Joint National Commission on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC) have considered “normal” blood pressure to be what was formerly considered “optimal” blood pressure, and a new category known as “prehypertension” was added. The prehypertension category describes a range of blood pressure measurements that are border-line of hypertension classification. The blood pressure measurements in the prehypertension category are systolic measurements from 120-139 mmHg and diastolic measurements from 80-89 mmHg (Liszka et. al, 2005).

A report from the Framingham Heart Study, a prospective study of the development of cardiovascular disease among adult US men and women, found that individuals with blood pressure within the “prehypertension” range have double the risk of cardiovascular disease compared to those with blood pressure in the optimal range (Miller & Jehn, 2004). Similarly, evidence suggests that decreasing systolic blood pressure by 20 mmHg and diastolic blood pressure by 10 mmHg in a middle-aged population with no prior history of cardiovascular disease reduces the risk of cardiovascular mortality by one half (Liszka et. al, 2005).

The prehypertension category has great potential to save lives by increasing awareness for the need for lifestyle modification. However, despite the potential for early lifestyle modification, this may not always be practical. One concern regarding the expectation that prehypertension classification would result in lifestyle modification is that it is often difficult for patients to alter their lifestyles. Patients are likely to be reluctant to change their diet or to increase physical activity and, therefore, the expectation that they would willingly abide by these recommendations may not be realistic. Another issue is that some patients with blood pressures

in the prehypertension range may not have access to health care services due to lack of health insurance. This results in uninsured prehypertensive patients being much less aware of their risk for hypertension than insured prehypertensive patients (Ayanian et. al, 2003). Ultimately, this introduces class, and possibly racial, disparities in awareness of hypertension risk. This is an area for public health intervention.

Despite its implications for predicting hypertension risk and encouraging lifestyle modifications, the prehypertension classification will not be used in this study. The data set used for this study was collected during the period from 1988-1994, which is prior to the time in which the new classification guidelines were introduced. Thus, the application of new diagnostic criteria to data that predates the criteria would result in an inaccurate measure of hypertension diagnosis.

Purpose and Hypotheses

This thesis seeks to answer the question of whether there is a relationship between hypertension diagnosis and smoking cessation. Specifically, this thesis seeks to answer the following question: Does lack of knowledge of hypertension status ultimately lead to continual smoking behavior, reducing the chance of cessation? A hypothesis related to this question is that among those that have attempted to quit smoking, those with undiagnosed hypertension will be less likely to successfully quit.

One theory that supports this hypothesis is the Health Belief Model. An assertion of this theory is that an individual will adopt a health-related behavior change only if the individual perceives that a negative health condition may be avoided (Cronin, 1986). In regards to this thesis, the health belief model supports the hypothesis that those with diagnosed hypertension are

more likely to sustain smoking cessation because they perceive that the negative health effects of hypertension (such as heart failure and kidney failure) may be avoided. Conversely, those with undiagnosed hypertension are unaware that they have the condition and, therefore, less likely to abstain because they do not foresee the negative consequences of continuing their smoking behavior.

Another theory relevant to this thesis is the Theory of Planned Behavior/Reasoned Action, which relates behavior change to an individual's *intention* to change their behavior. This theory defines 'intention' as a measure of a person's *readiness* to perform the behavior (Galvin, 1992). Individuals with undiagnosed hypertension are less inclined to have the intention to quit smoking (as explained by the Health Belief Model) and therefore, they are less likely to be successful at modifying their health behavior and lifestyle.

A second question this thesis seeks to address is whether the relationship between smoking cessation and hypertension diagnosis will be consistent regardless of the statistical methods used in the analyses. Specifically, the analyses will be conducted in two ways: one way using an "biostatistical" approach and the other using an "epidemiological" approach. The epidemiological approach places emphasis on the clinical significance of variables in the regression model and categorizes all continuous variables on the basis of this clinical significance. The biostatistical approach emphasizes statistical significance and appropriate scales for continuous variables in the regression model. If both regression models indicate the same relationship between smoking cessation and hypertension diagnosis, this would provide sufficient support that such a relationship exists.

Gaps in Current Knowledge

There are several gaps in knowledge regarding the relationship between smoking and hypertension. For example, an explanation of whether smoking leads to the development of hypertension is still unclear, and there is not enough evidence linking smoking as a causative factor for hypertension. In addition, there is virtually no data on long-term trends in smoking in hypertensive patients at the population level (Kastarinen et. al, 2002).

One finding that helps fill some of the gaps is that smoking causes an acute rise in blood pressure by contributing to atherosclerotic processes (Niskanen et. al, 2004). Exactly how this mechanism works remains to be revealed. However, a significant finding is that the cadmium found in cigarettes could potentially play a role, for the cadmium concentration in the kidneys of smokers is approximately double that of nonsmokers (Pardell et. al, 1998). Although not tested in human subjects, cadmium has been found to produce blood pressure elevations in rats. More research is needed on the potential role of cadmium as an intermediary in the link between smoking and hypertension.

Another puzzling conclusion is that smoking during pregnancy reduces the risk of both gestational hypertension and preeclampsia (Zhang et.al, 1999, England et. al, 2002). The reduced risk persisted even after the female stopped smoking. Surprisingly, the more and the longer a woman had smoked previously, the lower the risk of developing hypertension during pregnancy (Zhang et.al, 1999). Future studies on this puzzling association are needed.

To further complicate matters, where evidence does exist to explain the link between smoking and hypertension, there is conflicting data. Namely, there are divergent results on whether smokers have higher blood pressure than nonsmokers. The traditional accepted view has been that smokers show lower blood pressure than nonsmokers, yet several studies have

reported the contrary (Pardell et. al, 1998). Many authors have suggested that the reason smokers may have lower blood pressure than nonsmokers could be that smokers may generally have lower bodyweight than nonsmokers (Pardell et. al, 1998), which would decrease the risk of hypertension. Another possibility is that there could be an adaptation or “rebound” effect by which smokers may have blood pressure levels lower than nonsmokers during smoking-free intervals (Pardell et. al, 1998). An important consideration is whether smokers tend to quit after a diagnosis of hypertension is made, which could explain why diagnosis would be more frequent in ex-smokers (who are technically non-smokers) than current smokers (Gulliford, 2001). This thesis seeks to explore this possibility.

It is important to consider the reason for inconsistent conclusions. One area for discrepancy may be the potential for errors in measuring blood pressure. It is not uncommon to obtain inaccurate readings as a result of inaccurate cuff size or natural blood pressure fluctuations due to exercise, stress, relaxation, the time of day, alcohol intake, caffeine, or medications (Getliffe et. al, 2000). Thus, it is recommended to perform *repeated* measurements and report the average of these measurements, but some studies fail to follow this recommended procedure.

Similarly, it is often difficult to accurately measure smoking cessation. For example, some studies measure smoking cessation in terms of “ever quitting”. Thus, individuals are considered “ex-smokers” based on self-reports that they have quit smoking, yet there is often not a consideration of how long they have been abstinent. The recommended approach for accurately reflecting smoking cessation is to measure “duration of smoking” as an individual’s age at quitting minus their reported age at initiation (McGrady & Pederson, 2002). Then, the

successful quit ratio may be determined for this population by calculating the proportion who have quit for at least 12 months (King et.al, 1999).

This thesis will help to fill some of the gaps in knowledge of the relationship between smoking and hypertension. One possible finding is a resolution of the paradox of whether hypertension diagnosis leads to smoking cessation or if a subconscious fear of developing hypertension, manifested as a fear of weight gain, prevents smoking cessation. By helping to fill the gaps, this thesis would advance public health knowledge in two leading dilemmas threatening the nation's health. To the best of my knowledge, no other paper has attempted to thoroughly explore the relationship between hypertension diagnosis and smoking cessation, and this is the first paper to suggest undiagnosed hypertension as a barrier to smoking cessation.

Chapter II: Review of Literature

Evidence for the Link Between Smoking and Hypertension

There is much evidence suggesting a causal relationship between smoking and hypertension. One study revealed that tobacco smoking has interactive effects with hypertension, resulting in increased risk for chronic diseases such as coronary heart disease (Schmitz et. al, 1999). Smoking is also believed to be associated increasing the risk factors of the metabolic syndrome, a group of metabolic disorders which include abdominal obesity, dyslipidemia, and hypertension (Tonstad & Mette, 2005). Furthermore, smoking raises the risk of stroke in hypertensive men. According to a Japanese study of hypertensive men, the population attributable fraction associated with current smoking was 28% for total stroke, 33% for coronary heart disease, and 31% for total cardiovascular disease (Yamagishi et. al, 2003).

In addition to personal increased cardiovascular disease risk, smokers also increase the risk for others. That is, exposure to environmental tobacco smoke results in an elevated risk of cardiovascular disease among nonsmoking subjects, and occupational exposure to second hand smoke could increase the risk of developing acute coronary syndromes (Panagiotakos et. al, 2002). In light of this new finding, it is important for the tobacco control community to increase efforts of protecting nonsmokers from the health consequences of secondhand smoke.

Evidence for the Interaction of Smoking and Hypertension

The presence of both smoking and hypertension present competing risks for a given individual. Specifically, smoking and hypertension may present an additive increased risk

for cardiovascular disorders. For example, one study found that smoking and hypertension interactively play a key role in the initiation of atherosclerosis by generating endothelial dysfunction and vascular damage (Mazzone et. al, 2001). Thus, when hypertension and smoking coexist in the same individual, the final outcome is a significantly increased risk of cardiovascular diseases.

Consequently, hypertensive patients who smoke may ultimately face premature mortality as a result of the interactive risks of cigarette smoking and persistently high blood pressure among this group. This is confirmed by another study which found that the total cardiovascular risk for hypertensive patients who smoke is much greater compared to their nonsmoking counterparts (Kastarinen et. al, 2002). The study concluded that smoking is one of the main determinants for excess mortality among hypertensive patients. This reinforces the importance of implementing smoking cessation strategies targeting hypertensive patients.

Evidence For Inequalities in Smoking and Hypertension

Gender, class, and racial inequalities are evident in adult smoking and hypertension rates. In regards to smoking, some evidence exists which suggests that women are more likely than men to smoke in response to psychological stress (Schmitz, Spiga et al. 1999). Other evidence suggests that men are more likely to quit than women (McGrady and Pederson, 2002). With respect to this difference in cessation, it is possible that health beliefs may influence women's cessation. Specifically, considering that many anti-smoking ads target men, women may perceive themselves as less vulnerable to developing smoking-related health problems (Schmitz, Spiga et al. 1999).

Gender differences in hypertension also exist. During 1999-2002, the prevalence of hypertension was higher among women than men, and men were more likely to have undiagnosed hypertension (Centers for Disease Control and Prevention 2005). The reason for gender differences in blood pressure levels is not yet completely understood, but sex hormones have been discussed as a possible cause (Os, Oparil et al. 2004).

Class inequalities are evident in smoking behavior. Smoking prevalence tends to be highest among those with less education, less income, and in occupations classified as “blue collar” jobs (Barbeau et. al, 2004). One possible explanation for the higher smoking prevalence of blue collar workers is that blue collar workers are less likely to be employed at worksites with restrictions on smoking, that is, they are more likely to work in tobacco-friendly environments (Honjo et. al, 2006). With respect to smoking cessation, class inequalities exist in that members of lower social classes are less likely to quit (Lawlor et.al, 2003 & Kiefe et. al, 2001) and may also have fewer resources for smoking cessation. Specifically, smokers in high social classes are more likely to use nicotine replacement therapy in their quit attempts compared with smokers in lower social classes (Honjo et. al, 2006).

Evidence of racial/ethnic disparities in smoking and hypertension are well-documented. In 2004, Hispanics and non-Hispanic Blacks had a lower prevalence of current smoking (15% and 20, respectively) than non-Hispanic Whites (22%) (Centers for Disease Control and Prevention, 2005). As for smoking cessation, Whites are more likely to quit than Hispanics and non-Hispanic Blacks (McGrady & Pederson, 2002) and more likely to receive smoking cessation advice from a health care provider (Houston et. al, 2005).

Although certain factors such as social norms, attitudes, family and community pressures, and rejection of media influence may contribute to the lower prevalence of smoking among

Blacks in comparison to Whites (Gritz et. al, 2003), differences exist even *within* the Black population. An examination of differences in smoking prevalence among Black adults by state in the United States found that smoking prevalence for Blacks was higher (greater than 25%) in the Midwestern states than in most Southern states, suggesting the possibility of regional differences in socio-cultural influences on smoking behavior among Blacks (Polednak & King, 1999). Related to this finding is another revelation that native-born African Americans were more likely to be current smokers than foreign-born blacks (King et. al, 1999), which further demonstrates the diversity in smoking behavior among Blacks.

To suggest that racial differences exist only in smoking behavior and not in hypertension would be incorrect. During 1999-2002, the age-adjusted prevalence of hypertension was 41% among non-Hispanic Blacks, 27% among non-Hispanic whites, and 25% among Mexican Americans; the proportion aware of having hypertension was 70% among non-Hispanic blacks, 63% among non-Hispanic whites, and 50% among Mexican Americans (Centers for Disease Control and Prevention, 2005). Therefore, Blacks have a higher prevalence of diagnosed hypertension in comparison to Whites and Whites have a higher prevalence of undiagnosed hypertension in comparison to Blacks. One explanation for this finding is that abdominal obesity is closely associated with racial/ethnic differences in the risk of hypertension, and a having larger than normal waist girth is associated with 1.58 and 1.39 increased risk of hypertension in black men and black women, respectively, adjusting for confounders (Okosun et. al, 2001). Other racial differences in hypertension are that African Americans are often more likely to be non adherent to blood pressure medication and are also more likely to perceive high blood pressure as “serious” (Bosworth et. al, 2006).

Chapter III: Methods and Procedures

Data Source

The 1988-1994 National Health and Nutrition Examination Survey (NHANES) collected by the National Center for Health Statistics of the Center for Disease Control and Prevention (CDC) was used for this study. NHANES is an on-going survey collected by CDC to assess the health of adults and children in the United States. The data collection methods used by NHANES have been described in previous studies (Liszka et. al, 2005). Briefly, NHANES collects health and nutrition data from U.S. adults and households through the distribution of questionnaires and performing physical examinations (NCHS, 1991). This study used data from both the questionnaires and physical examinations.

NHANES data is available to the public from the National Center for Health Statistics section of the CDC website. The NHANES data contained three SAS files: Adult, Lab, and Exam. Using these files, three new SAS files were created consisting of the variables of interest for this thesis. As the variables of interest were selected, the variable labels were changed from the original NHANES variable labels to labels that could easily be interpreted. Next, the variables from the new SAS files were merged into one file (using the NHANES sequence number as the key variable) to create a merged data set. After merging the data files, the categorical variables in the dataset were recoded such that '0' represented a decreased risk factor and '1' represented an elevated risk factor. To achieve this recoding of the variables, appropriate "if-then" statements were written in the SAS files. For each variable, if there was an "unknown" or a "not applicable" response, the value was re-coded as a missing value.

Study population

The population was selected based on age, sex, race, smoking cessation status, and hypertension diagnosis status. Only individuals with non-missing responses for smoking cessation and diagnosis status were eligible for this study. In addition, this study was limited to adults who were 18 years old or older.

Outcome definitions

For this study, the key independent variable was smoking cessation, and the key dependent variable was hypertension diagnosis. In the NHANES data, smoking status was captured by asking the individuals if they had ever smoked 100 or more cigarettes (equivalent to 5 packs of cigarettes) in their lifetime and if they currently smoked cigarettes. For the purpose of this study, individuals responding ‘yes’ to both questions were considered to be current smokers.

Previous studies have considered successful smoking cessation as remaining abstinent for one year or longer (McGrady & Pederson, 2002). Due to limitations in the data set, the same definition could not be used for this study. For this study, successful smoking cessation was defined as answering ‘yes’ to the question ‘have you ever smoked 100+ cigarettes in your lifetime’ and answering ‘no’ to the question ‘do you smoke now’.

For this analysis, hypertension was defined as having a diastolic blood pressure reading greater than or equal to 90 mm Hg, a systolic blood pressure reading greater than or equal to 140 mm Hg or current treatment with prescribed anti-hypertension medication. Hypertension diagnosis was determined on the basis of self-report of diagnosis by a doctor. NHANES provided repeated systolic and diastolic readings for each individual and determined the average of these readings. The average systolic and average diastolic readings were the variables selected for this study.

For the purpose of describing the study population, a binary variable for hypertension status was created such that individuals with a systolic blood pressure reading greater than or equal to 140mmHg and a diastolic reading greater than or equal to 90 mmHg or reported taking anti-hypertension medication were considered hypertensive and all others not meeting this criterion were considered not to have hypertension. Similarly, a binary hypertension diagnosis variable was created such that those who reported being told by a physician as having hypertension were considered to have been diagnosed and all others were considered to not have been diagnosed. This formed the key dependent variable and was used to assess the odds of prior hypertension diagnosis that was associated with each of the other variables.

For the purpose of describing the population in terms of the number of hypertensive individuals who were diagnosed and those who were undiagnosed, a second binary hypertension diagnosis variable was created. For this variable, those who were met this study's definition of hypertension and who reported that they had been diagnosed by a physician were considered to have diagnosed hypertension. Those reporting that they had not been told by a doctor that they were hypertensive, yet met the criteria for hypertension as defined by this study, were considered to have undiagnosed hypertension.

Potential confounders

Potential confounders in this study included demographic, biological, and lifestyle characteristics of the study population. These were defined based on the 1988-1994 NHANES definitions, and also based on definitions from the National Library of Medicine (NLM). The demographic variables used in this study included gender, age, race/ethnicity, and educational level.

NHANES data contains four race/ethnicity categories: non-Hispanic White (White), non-Hispanic Black (Black), Mexican-American, and Other. For all analyses, “dummy” variables were created for the racial groups. Specifically, a race1 variable was created such that if the race was ‘White’, then race1=1 and race1=0 otherwise. Similar race2 and race3 variables were created for the Black and Mexican-American races, respectively.

When performing the analyses using the epidemiological approach, ages of individuals were categorized into three groups: 18-39, 40-59, and 60+ years, representing increasing levels of hypertension risk according to the National Center for Health Statistics’ chart book on US trends in health (NCHS, 2005). Dummy variables were created such that age2=1 if the age group was 40-59 years (0 otherwise) and age3=1 if the age group was 60+ years (0 otherwise). Education was categorized into two levels: high school or less and not less than high school. When performing the analyses using the biostatistical approach, the age and education variables remained continuous variables.

The lifestyle variables used in this study were measures of cardiovascular physical activity (jogging/running and aerobics) and alcohol consumption. NHANES asked individuals whether they had jogged or ran in the past month and whether they had done aerobics in the past month. Individuals were considered to be physically active if they answered "yes" to this question and physically inactive if they answered "no" to the question. To measure alcohol consumption, NHANES asked if the individual had ever had 12 drinks in a lifetime. NHANES defined one drink as 12 ounces of beer, four ounces of wine, or one ounce of liquor. Individuals who responded that they had ever had 12 drinks in a lifetime were considered to be drinkers.

For biological and anthropometric variables, HDL cholesterol and body mass index (BMI), were used. Definitions from the National Library of Medicine were used in defining these

variables. A normal HDL cholesterol level was considered to be greater or equal to 1mmol/L, and a low HDL level was less than 1mmol/L. Individuals with a BMI value of less than 25 kg/m² were considered normal, a BMI of 25-30 kg/m² was considered overweight, and a BMI of greater than 30 kg/m² was considered obese. Using the statistical approach, the HDL and BMI variables remained continuous variables.

Statistical Analyses: Biostatistical Method

All statistical analyses were performed using SAS versions 8.2 and 9.1.

Step 1: Describing the study population

To describe the study population, independent t-tests and Pearson chi-square statistics were performed to assess the differences across hypertension diagnosis status for continuous and categorical variables, respectively. In addition, the frequency counts of categorical variables and the mean of continuous variables were determined. Continuous variables were also assessed for normality and collinearity with other continuous variables.

Step 2: Univariable analyses

Univariable logistic regression analyses were performed for categorical and continuous variables to determine their significance to the dependent variable at the 0.25 level of significance. Variables significant at the 0.25 level were considered candidates for the final multivariable model.

Step 3: Model-building

a) Fitting a multivariable model containing terms significant at the .25 level of significance

The model-building process began with fitting a multivariable model containing the variables significant at the 0.25 level in the univariable analyses. Variables not significant in this preliminary main effects model were removed from the analyses, forming the main effects model.

b) Checking linearity in the logit for continuous variables

The continuous variables were checked for the linearity in the logit assumption by examining smoothed scatter plots of the percent of subjects diagnosed with hypertension vs. the continuous variable. Then, the fractional polynomial transformation technique was used in order to “fix” nonlinearity in the logit. Briefly, this technique selects appropriate powers for the continuous variables using maximum log likelihood estimates.

c) Checking for interactions

After finding the appropriate scale for continuous variables, the next step was to check for possible interactions among the candidates eligible for inclusion in the final multivariate model. Interaction terms were added one-at-a-time to the main effects model and all terms significant at the .05 level of significance were considered candidates for the final multivariate model. After the interactions check, significant interaction terms were added to the main effects model to further assess their significance. Insignificant interaction terms were removed from the model.

d) Applying stepwise regression to select the best model

To select the final multivariate model, stepwise regression was performed for non-interaction terms and interaction terms separately. For the non-interaction terms, original continuous variables were included in the model as opposed to the fractional polynomial term(s). Only if the continuous variable was selected by the stepwise regression method was the fractional polynomial term(s) added to the model. In selecting the best model from among the variables selected by the stepwise regression process, the model with the smallest AIC value, the highest likelihood ratio, and with significant Wald chi-square values for all parameters was selected. Best subsets regression was also performed as an alternative to the stepwise selection

method, but the candidate models from this selection method did not have significant parameter estimates, so the stepwise regression method was used as the final selection criterion.

Step 4: Assessing the fit of the final model and interpretations

The next step was to assess the fit and diagnostics of the model selected from the stepwise regression method. The Hosmer-Lemeshow and Osiris and Rojek goodness of fit tests were performed to assess the fit of the model. In addition, the area under the ROC curve, classification tables, and Pearson chi-square and deviance summary statistics also helped to assess the fit. Model diagnostics were performed using three plots: delta chi-square vs. estimated probabilities, delta deviance vs. estimated probabilities, and delta beta vs. estimated probabilities. Finally, interpretations of the model were performed by examining the estimated coefficients, standard errors, 95% confidence intervals, and odds ratios for the parameter estimates. Interaction terms required separate odds ratio analyses.

Statistical Analyses: Epidemiological Method

The epidemiological approach was performed in much the same way as the biostatistical approach. The only difference is that the continuous variables were categorized on the basis of clinical significance. Therefore, many of the procedures used for continuous variables in the biostatistical approach, such as fractional polynomial transformations, were omitted using this approach. The details on how the continuous variables were categorized can be found in the SAS codes in the Appendix pages 92. Specifically, the age variable that was continuous in the biostatistical analyses was categorized into three groups: 20-39 years, 40-59 years, and 60+ years. Similarly, the education variable was categorized into two groups: high-school graduate or less vs. not high-school graduate or less; the systolic blood pressure variable was categorized into two groups: blood pressure greater than or equal to 120 mmHg

(hypertensive) vs. blood pressure not greater than or equal to 120 mmHg (non-hypertensive); the diastolic blood pressure variable was categorized into two groups: blood pressure greater than or equal to 90 mmHg (hypertensive) vs. blood pressure not greater than or equal to 90 mmHg (non-hypertensive); the HDL cholesterol variable was categorized into two groups: HDL level greater than or equal to 1 (normal HDL level) vs. hdl level not greater than or equal to one (abnormal); and the BMI variable was categorized into two groups: BMI greater than or equal to 30 (obese) and BMI not greater than or equal to 30 (non obese).

Chapter IV: Results

Descriptive characteristics

Tables 1 and 2 (pages 25-26) show the characteristics of the study population. The final study population was comprised of 17,589 males and females (8,175 and 9,414 respectively) age 20 – 90 years. The mean age was 44 years. The eligible population included 40.54% Non-Hispanic Whites, 27.51% Non-Hispanic Blacks, and 29.54% Mexican-Americans. In addition, the mean education level in the study population was 12.31 years of school completed and the mean BMI was 26.41.

Of the total eligible sample, 25.37% of subjects were smokers and 74.63% were non-smokers. In terms of smoking cessation, 23.69 % of the study population consisted of former smokers who had successfully quit and 74.13% of the study population had failed to quit. In regard to hypertension, 29.89% of the population was hypertensive and 67.93 % were not hypertensive.

Table 3 (page 26) and Table 4 (page 27) present the results of the check for normality and collinearity, respectively, for the continuous variables. The normality assumption was checked by examining the mean and median of each distribution and determining if there were extreme values for skewness and kurtosis. Variables with the mean approximately equal to the median and with non-extreme values for skewness and kurtosis were considered to be normal.

According to Table 3, the continuous variables in the dataset satisfied the normality assumption. To check for collinearity, a linear regression analysis was performed in which each continuous variable was regressed on all other continuous variables and the variance inflation factors (VIFs)

Table 1: Frequency of Categorical Variables N=17,589

| Risk Factor | Variable | N | % Study Population |
|----------------------------------|--------------------------|----------|---------------------------|
| Race | Non-Hispanic White | 7290.00 | 41.45 |
| | Non-Hispanic Black | 4946.00 | 28.12 |
| | Mexican-American | 4664.00 | 26.52 |
| | Other | 689.00 | 3.92 |
| Gender | Male | 8175.00 | 46.48 |
| | Female | 9414.00 | 53.52 |
| Age | Less than 20 yrs old | 705.00 | 4.01 |
| | 20-39 | 6752.00 | 38.39 |
| | 40-59 | 4433.00 | 25.20 |
| | 60 + | 5699.00 | 32.40 |
| Physical Activity Level | Physical activity | 2027.00 | 11.52 |
| | No physical activity | 15562.00 | 88.48 |
| Alcohol Use | Non-drinker | 10085.00 | 57.34 |
| | Drinker | 7504.00 | 42.66 |
| Smoking Status | Non-smoker | 13127.00 | 74.63 |
| | Smoker | 4462.00 | 25.37 |
| Smoking Cessation | Successful | 4259.00 | 24.21 |
| | Unsuccessful | 13330.00 | 75.79 |
| Hypertension Status | Non-hypertensive | 12215.00 | 69.45 |
| | Hypertensive | 5374.00 | 30.55 |
| Knowledge of Hypertension | Undiagnosed Hypertension | 1750.00 | 32.56 |
| | Diagnosed Hypertension | 3624.00 | 67.44 |

Table 2: Mean of Continuous Variables

| Variable | Mean |
|--------------------------|-------------|
| Education | 12.311863 |
| Age | 43.837959 |
| Systolic Blood Pressure | 122.082003 |
| Diastolic Blood Pressure | 73.904431 |
| Body Mass Index | 26.406748 |
| HDL cholesterol | 1.310567 |

Table 3: Normality Check for Continuous Variables

| Variable | Mean | Median | Skewness | Kurtosis |
|-----------------|-------------|---------------|-----------------|-----------------|
| Education | 10.83 | 12.00 | -0.82 | 0.49 |
| Age | 48.33 | 45.00 | 0.30 | -1.44 |
| Systolic | 126.43 | 122.00 | 1.00 | 1.29 |
| Diastolic | 74.18 | 74.00 | 0.220 | 1.30 |
| HDL cholesterol | 1.32 | 1.27 | 1.21 | 3.88 |
| Body Mass Index | 26.98 | 26.10 | 1.24 | 3.20 |

Table 4: Collinearity Check for Continuous Variables

Education as the dependent variable

| Obs | Variable | Variance Inflation |
|-----|-----------|-----------------------|
| 1 | Intercept | 0 |
| 2 | AGE | 1.65505 |
| 3 | systolic | 2.16959 |
| 4 | diastolic | 1.44191 |
| 5 | hdlcholes | 1.06055 |
| 6 | BMI | 1.11990 |

Age as the dependent variable

| Obs | Variable | Variance Inflation |
|-----|-----------|-----------------------|
| 1 | Intercept | 0 |
| 2 | education | 1.05453 |
| 3 | systolic | 1.40704 |
| 4 | diastolic | 1.39657 |
| 5 | hdlcholes | 1.06334 |
| 6 | BMI | 1.12073 |

Systolic as the dependent variable

| Obs | Variable | Variance Inflation |
|-----|-----------|-----------------------|
| 1 | Intercept | 0 |
| 2 | AGE | 1.08070 |
| 3 | education | 1.06176 |
| 4 | diastolic | 1.08056 |
| 5 | hdlcholes | 1.06286 |
| 6 | BMI | 1.11693 |

Diastolic as the dependent variable

| Obs | Variable | Variance Inflation |
|-----|-----------|-----------------------|
| 1 | Intercept | 0 |
| 2 | AGE | 1.61472 |
| 3 | education | 1.06224 |
| 4 | systolic | 1.62661 |
| 5 | hdlcholes | 1.06262 |
| 6 | BMI | 1.09515 |

Hdlcholes as the dependent variable

| Obs | Variable | Variance Inflation |
|-----|-----------|-----------------------|
| 1 | Intercept | 0 |
| 2 | AGE | 1.68161 |
| 3 | education | 1.06866 |
| 4 | systolic | 2.18843 |
| 5 | diastolic | 1.45345 |
| 6 | BMI | 1.06383 |

BMI as the dependent variable

| Obs | Variable | Variance Inflation |
|-----|-----------|-----------------------|
| 1 | Intercept | 0 |
| 2 | AGE | 1.68040 |
| 3 | education | 1.06989 |
| 4 | systolic | 2.18041 |
| 5 | diastolic | 1.42021 |
| 6 | hdlcholes | 1.00863 |

were observed. Any variable with a VIF of 10 or greater was considered to be collinear with the dependent variable. According to the results, there was no strong evidence of collinearity.

Hypothesis Testing

Tables 5 and 6 (below) present the results of testing the associations of categorical and continuous variables, respectively, to the independent variable. The Pearson chi-square test was used to determine if an association existed between the categorical variables and dependent variable, and the Cramer's V statistic measured the strength of the association. The results indicate that all categorical variables were associated with the dependent variable, but the association was not very strong (the Cramer's V statistic was not greater than or equal to 0.5). T-tests were used to test the association of the continuous variables to the dependent variable. The results show that all continuous variables are associated with the dependent variable.

Table 5: Testing the Association of the Categorical Variables to the Dependent Variable

| Variable | $\chi^2_{(1)}$ | p-value | Cramer's V statistic |
|-------------------|----------------|---------|----------------------|
| White | 59.68 | <.0001 | .0554 |
| Black | 54.67 | <.0001 | .0530 |
| Mexican-American | 215.40 | <.0001 | -.1053 |
| gender | 62.81 | <.0001 | .0569 |
| Physical activity | 307.11 | <.0001 | .1257 |
| Alcohol | 270.73 | <.0001 | .1181 |
| Cessation | 167.90 | <.0001 | -.0930 |

Table 6: Testing the Association of the Continuous Variables to the Dependent Variable

| Variable | t-score | p-value |
|-----------------|---------|---------|
| Education | 11.52 | <.0001 |
| Age | -49.20 | <.0001 |
| Systolic | -68.66 | <.0001 |
| Diastolic | -39.14 | <.0001 |
| HDL cholesterol | 4.78 | <.0001 |
| Body Mass Index | -30.03 | <.0001 |

The results of the univariable analyses are shown in Table 7 (below). Results indicate that all categorical and continuous variables are significant to the model at the 0.25 level of significance. Table 8 (below) shows the results of fitting a preliminary main effects model using the variables significant at the 0.25 level. According to the results, education is not significant to the model and, therefore, is removed from further consideration for contributing to a multivariate model.

Table 7: Univariable Analyses

| Variable | Wald Chi-Square | p-value |
|-------------------|-----------------|---------|
| Gender | 62.67 | <.0001 |
| White | 59.56 | <.0001 |
| Black | 54.52 | <.0001 |
| Mexican-American | 211.87 | <.0001 |
| Alcohol | 211.06 | <.0001 |
| Physical Activity | 281.10 | <.0001 |
| Cessation | 166.45 | <.0001 |
| Age | 1960.46 | <.0001 |
| Education | 130.69 | <.0001 |
| Systolic | 2891.66 | <.0001 |
| Body Mass Index | 776.22 | <.0001 |
| Diastolic | 1287.35 | <.0001 |
| HDL cholesterol | 22.73 | <.0001 |

Table 8: Fitting a Preliminary Main Effects Model

| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---------------------------------------|----|----------|----------------|-----------------|------------|
| Intercept | 1 | -4.9733 | 0.2562 | 376.7849 | <.0001 |
| gender Male vs. Female | 1 | -0.1119 | 0.0500 | 5.0159 | 0.0251 |
| race1 White vs. Not-White | 1 | 0.0711 | 0.1323 | 0.2885 | 0.5912 |
| race2 Not Black vs. Black | 1 | -0.5642 | 0.1343 | 17.6629 | <.0001 |
| race3 Not M-A vs. M-A | 1 | 0.1186 | 0.1369 | 0.7496 | 0.3866 |
| alcohol Non-drinker vs. Drinker | 1 | 0.0971 | 0.0477 | 4.1433 | 0.0418 |
| pajog Phys. Act. vs. No Phys. Act. | 1 | -0.2344 | 0.0882 | 7.0600 | 0.0079 |
| cessation Unsuccessful vs. Successful | 1 | -0.0441 | 0.0500 | 0.7780 | 0.3777 |
| AGE | 1 | 0.0411 | 0.00146 | 789.9175 | <.0001 |
| education | 1 | -0.00204 | 0.00666 | 0.0936 | 0.7596 |

| | | | | | |
|-----------|---|---------|---------|----------|--------|
| BMI | 1 | 0.0864 | 0.00406 | 452.9234 | <.0001 |
| hdlcholes | 1 | -0.1000 | 0.0607 | 2.7152 | 0.0994 |

In testing linearity in the logit for continuous variables, smoothed scatter plots were constructed for the percent of subjects diagnosed vs. the continuous variable (Figures 1-5, pages30-32).

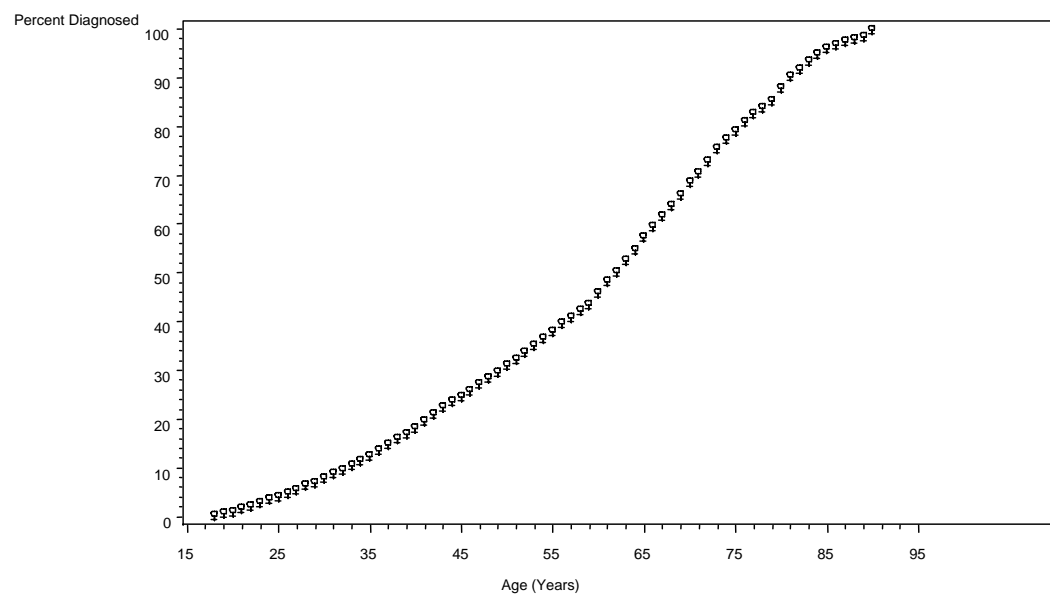


Figure 1: Smoothed Scatter plot of Patients Diagnosed vs. Age

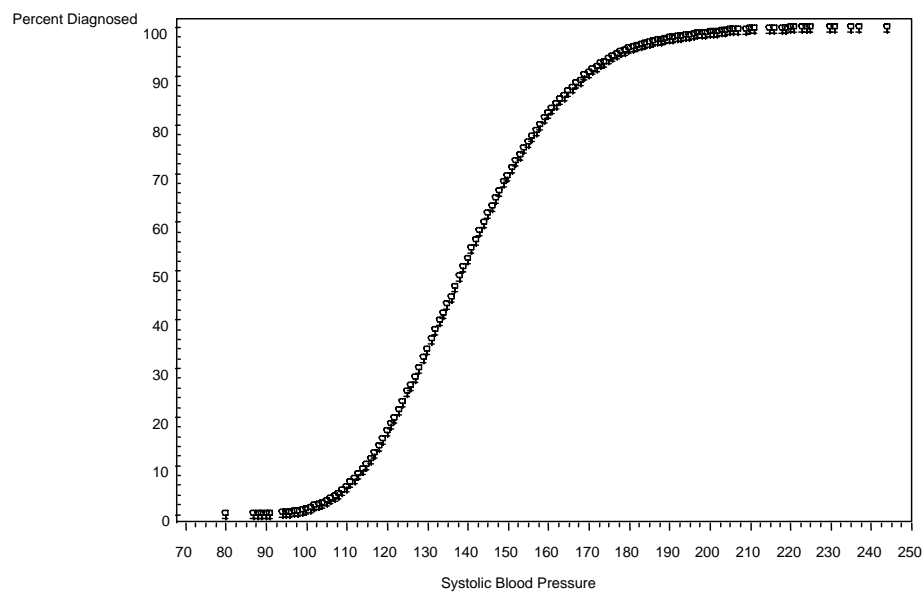


Figure 2: Smoothed Scatter plot of Patients Diagnosed vs. Systolic Blood Pressure

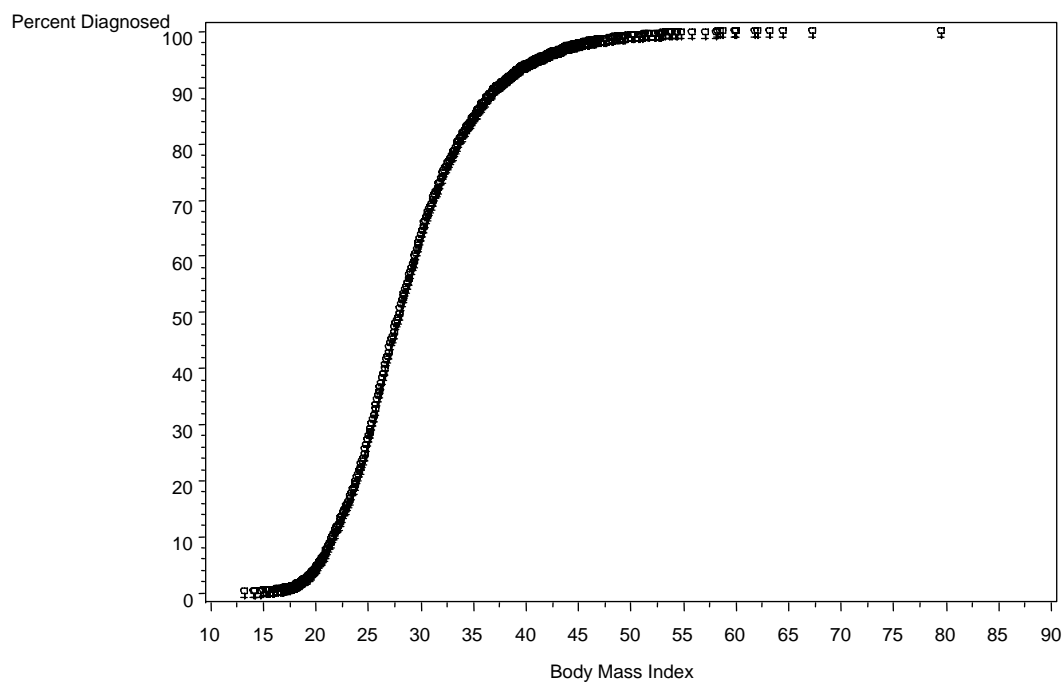


Figure 3: Smoothed Scatter plot of Patients Diagnosed vs. BMI

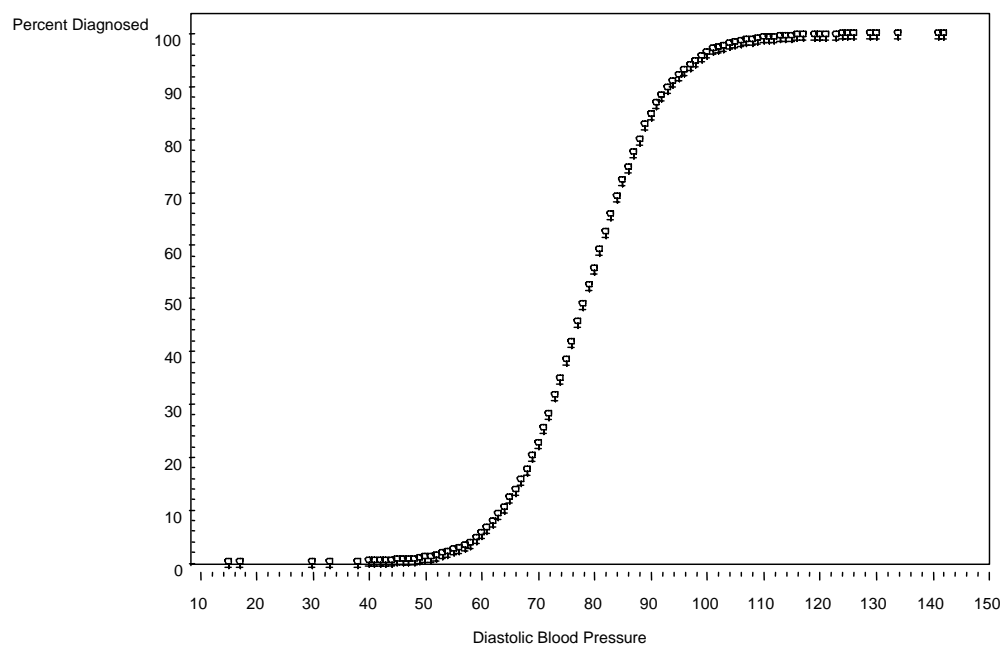


Figure 4: Smoothed Scatter plot of Patients Diagnosed vs. Diastolic Blood Pressure

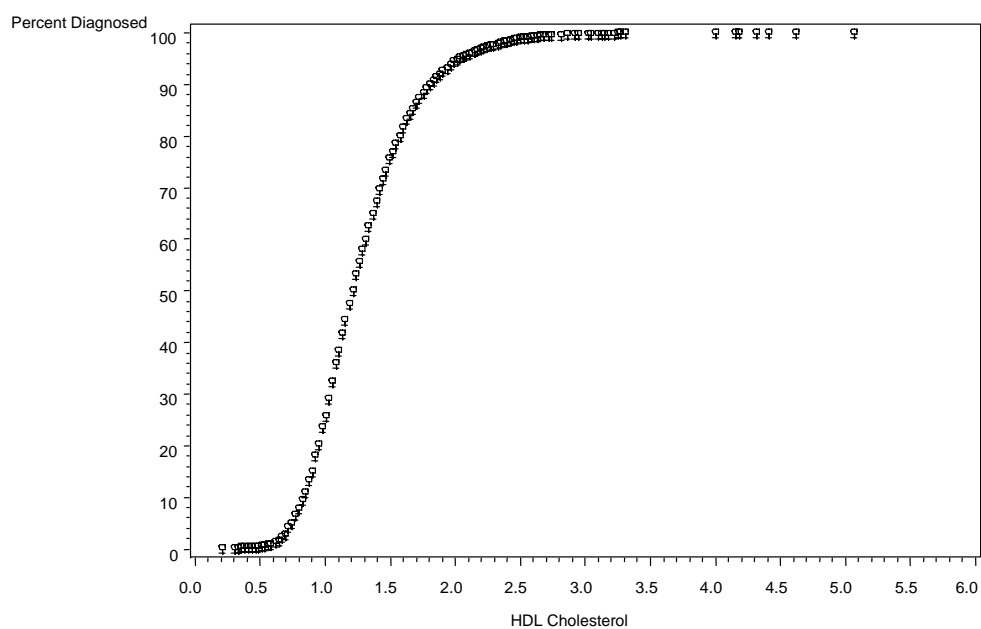


Figure 5: Smoothed Scatter plot of Patients Diagnosed vs. HDL cholesterol

All continuous variables seemed to have problems with linearity in the logit, so the fractional polynomial transformation technique was performed for all variables to determine the appropriate powers of the continuous variable. See the Appendix pages 70-80 for the SAS codes used to perform the fractional polynomial transformation technique. Table 9 (page 33) summarizes the results of applying the fractional polynomial techniques. The results indicate that the age, bmi, hdl cholesterol and education variables all needed transformations. Transformation variables were as follows: $\text{age1} = \text{age}^{-0.5}$, $\text{bmi1} = \ln(\text{bmi})$, $\text{hdlcholes1} = \text{hdlcholes}^{-1}$, $\text{education1} = \ln(\text{education})$, and $\text{education2} = \text{education}$.

To test for the presence of interactions, interaction terms were added one-at a time to the preliminary main effects model. Table 10 (page 34) illustrates the interaction terms significant at the .05 level of significance. These terms were added to the preliminary main effects model to

Table 9: Results of the Fractional Polynomial Transformation

| Variable: Age | | | | | | | |
|----------------------------------|----------|------------------|---------------------------|---------------------------------|----------------------|-------------------------------|-------------|
| | d.f. | Deviance | G for model Vs. Linear | p-value for model vs. Linear | G for J=2 vs. J=1 | p-value for J=2 vs. J=1 | powers |
| Not in Model | 0 | 13323.935 | _____ | _____ | _____ | _____ | _____ |
| Linear | 1 | 12458.567 | 0.000 | <.0001 | _____ | _____ | 1 |
| J=1 | 2 | 12442.995 | 15.572 | <.0001 | _____ | _____ | -0.5 |
| J=2 | 4 | 12439.213 | 19.354 | .0002 | 3.782 | 0.1509 | -0.5, 3 |
| Variable: Body Mass Index | | | | | | | |
| | d.f. | Deviance | G for model Vs. Linear | p-value for model vs. Linear | G for J=2 vs. J=1 | p-value for J=2 vs. J=1 | powers |
| Not in Model | 0 | 12954.883 | _____ | _____ | _____ | _____ | _____ |
| Linear | 1 | 12458.567 | 0.000 | <.0001 | _____ | _____ | 1 |
| J=1 | 2 | 12452.077 | 6.490 | .0108 | _____ | _____ | 0 |

| | | | | | | | |
|----------------------------------|----------|------------------|---------------------------|---------------------------------|----------------------|-------------------------------|------------|
| J=2 | 4 | 12452.063 | 6.504 | .0895 | 0.014 | 0.9930 | 0, 3 |
| Variable: Education | | | | | | | |
| | d.f. | Deviance | G for model Vs. Linear | p-value for model vs. Linear | G for J=2 vs. J=1 | p-value for J=2 vs. J=1 | powers |
| Not in Model | 0 | 12532.649 | _____ | _____ | _____ | _____ | _____ |
| Linear | 1 | 12458.567 | 0.000 | 0.7596 | _____ | _____ | 1 |
| J=1 | 2 | 12183.807 | 274.760 | <.0001 | _____ | _____ | 0 |
| J=2 | 4 | 12177.320 | 281.247 | <.0001 | 6.487 | 0.0390 | 0,1 |
| Variable: HDL cholesterol | | | | | | | |
| | d.f. | Deviance | G for model Vs. Linear | p-value for model vs. Linear | G for J=2 vs. J=1 | p-value for J=2 vs. J=1 | powers |
| Not in Model | 0 | 13208.772 | _____ | _____ | _____ | _____ | _____ |
| Linear | 1 | 12458.567 | 0.000 | 0.0994 | _____ | _____ | 1 |
| J=1 | 2 | 12454.488 | 4.079 | 0.0434 | _____ | _____ | -1 |
| J=2 | 4 | 12451.957 | 6.610 | 0.0854 | 2.531 | 0.2821 | -1, 3 |

Table 10: Interactions Significant at the .05 level

| Interaction term | p-value |
|-------------------------|----------------|
| bmi*race2 | .0002 |
| education1*race2 | .0011 |
| age1*race3 | .0342 |
| alcohol*pajog | .0228 |
| age1*alcohol | .0064 |
| hdlcholes*bmi1 | .0362 |
| age1*cessation | .0308 |
| education2*race1 | .0095 |
| race2*pajog | .0058 |
| age1*race2 | <.0001 |
| gender*race2 | .0056 |
| bmi1*gender | .0019 |
| education1*gende | <.0001 |
| race1*pajog | .0017 |
| age1*race1 | .0173 |
| bmi1*race1 | .0042 |
| gender*race3 | .0070 |

determine if they were significant at the .25 level in the multivariate model (Table 11, below).

The interaction term race1*bmi was removed from the model ($p > .25$).

Table 11: Fitting a Model Containing the Main Effects and Significant Interactions

| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|--|----|----------|----------------|-----------------|------------|
| Intercept | 1 | -1.0509 | 1.6309 | 0.4152 | 0.5193 |
| gender Male | 1 | -3.2693 | 0.8460 | 14.9319 | 0.0001 |
| race1 White | 1 | -2.0698 | 1.2971 | 2.5463 | 0.1106 |
| race2 Not Black vs. Black | 1 | -2.5146 | 1.3128 | 3.6690 | 0.0554 |
| race3 Not Mexican-American vs. Mexican American | 1 | 1.1824 | 0.8115 | 2.1233 | 0.1451 |
| alcohol Non-drinker vs. Drinker | 1 | 0.6780 | 0.2525 | 7.2114 | 0.0072 |
| pajog Physical activity vs. No physical Activity | 1 | -0.3668 | 0.1696 | 4.6760 | 0.0306 |
| cessation Unsuccessful vs. Successful | 1 | -0.8830 | 0.3008 | 8.6179 | 0.0033 |
| age1 | 1 | -24.8785 | 6.1324 | 16.4586 | <.0001 |
| bmi1 | 1 | 1.2587 | 0.4088 | 9.4783 | 0.0021 |
| hdlcholes1 | 1 | -1.4128 | 1.5683 | 0.8115 | 0.3677 |
| education1 | 1 | -0.1104 | 0.0248 | 19.7814 | <.0001 |
| education2 | 1 | 0.4261 | 0.1636 | 6.7834 | 0.0092 |
| bmi1*age2 Not Black | 1 | 0.5666 | 0.3217 | 3.1015 | 0.0782 |
| education1*age2 Not Black | 1 | 0.0193 | 0.0189 | 1.0474 | 0.3061 |
| age1*age3 Not Mexican-American | 1 | -7.5189 | 5.4896 | 1.8760 | 0.1708 |
| alcohol*pajog Non-drinker Physical activity | 1 | -0.3164 | 0.1957 | 2.6146 | 0.1059 |
| age1*alcohol Non-drinker | 1 | -3.6900 | 1.7418 | 4.4880 | 0.0341 |
| age1*cessation Unsuccessful | 1 | 6.0447 | 2.1617 | 7.8192 | 0.0052 |
| bmi1*hdlcholes1 | 1 | 0.4700 | 0.4709 | 0.9962 | 0.3182 |
| education2*age1 White | 1 | 0.1670 | 0.1670 | 0.9999 | 0.3173 |
| age2*pajog Not Black Physical activity | 1 | 0.2036 | 0.2402 | 0.7184 | 0.3967 |
| age1*age2 Not Black | 1 | -0.7514 | 5.5053 | 0.0186 | 0.8914 |
| gender*age2 Male Not Black | 1 | -0.1796 | 0.1097 | 2.6823 | 0.1015 |
| gender*age3 Male Not Mexican-American | 1 | 0.0966 | 0.1278 | 0.5716 | 0.4496 |
| bmi1*gender Male | 1 | 0.7876 | 0.2442 | 10.4047 | 0.0013 |
| education1*gender Male | 1 | 0.0511 | 0.0141 | 13.2120 | 0.0003 |
| age1*pajog White Physical activity | 1 | 0.3527 | 0.2272 | 2.4097 | 0.1206 |
| age1*age1 White | 1 | 5.7594 | 5.4019 | 1.1367 | 0.2863 |
| bmi1*age1 White | 1 | 0.2801 | 0.3060 | 0.8378 | 0.3600 |

Table 12 on page 36 shows the results of the stepwise regression model selection method.

The method was applied separately for interaction terms and non-interaction terms. Using this method, variables significant at the .15 level were considered for entry into the model and all variables significant at the .20 level remained in the model. According to the stepwise selection process, the cessation variable was not important to the model ($p > .25$).

The stepwise selection method closely followed the approach as outlined by Hosmer and Lemeshow in the second edition of Applied Logistic Regression. This method required the use of the original continuous variables (as opposed to the transformation variables) when selecting non-interaction terms. In order to verify the insignificance of the cessation variable in the model, the stepwise regression method was applied using the transformation variables instead of the original variables, deviating from the approach suggesting by Hosmer and Lemeshow. The

results were the same: the cessation variable was not significant to the model. As a final check of the significance of the cessation variable, the criteria for entry into the model was changed from .15 to .20 and the criteria for removal from the model was changed from .20 to .25. Again, the cessation variable was not significant.

Despite the different approaches in applying the stepwise selection method, no logistic regression model could be constructed using the cessation variable, and the question of whether smoking cessation may be explained by hypertension diagnosis could not be further investigated using the biostatistical approach. It is often the case that the association between response variables and regressors cannot be modeled directly due to their threshold relations. Even using transformations, their relation often still cannot achieve “linearity”. Therefore, the epidemiological approach was used as follows.

Table 12: Results of Stepwise Regression

Applying Stepwise Regression to non-interaction terms
Analysis of Effects Eligible for Entry

| Effect | DF | Score | |
|-----------|----|------------|------------|
| | | Chi-Square | Pr > ChiSq |
| race1 | 1 | 0.2973 | 0.5856 |
| cessation | 1 | 0.8017 | 0.3706 |
| education | 1 | 0.0506 | 0.8219 |

NOTE: No (additional) effects met the 0.15 significance level for entry into the model.

Summary of Stepwise Selection

| Step | Entered | Effect Removed | DF | Number In | Score Chi-Square | Wald Chi-Square | Pr > ChiSq |
|------|---------|-------------------|----|--------------|---------------------|--------------------|------------|
| 1 | AGE | | 1 | 1 | 1282.5903 | | <.0001 |
| 2 | BMI | | 1 | 2 | 658.0877 | | <.0001 |
| 3 | race2 | | 1 | 3 | 123.3297 | | <.0001 |
| 4 | pajog | | 1 | 4 | 9.4959 | | 0.0021 |

| | | | | | |
|---|-----------|---|---|--------|--------|
| 5 | race3 | 1 | 5 | 9.6206 | 0.0019 |
| 6 | alcohol | 1 | 6 | 8.1351 | 0.0043 |
| 7 | gender | 1 | 7 | 2.7171 | 0.0993 |
| 8 | hdlcholes | 1 | 8 | 2.6645 | 0.1026 |

Applying Stepwise Regression to interaction terms

| Step | Entered | Effect Removed | DF | Number In | Score Chi-Square | Wald Chi-Square | Pr > ChiSq |
|------|---------------|-------------------|----|--------------|---------------------|--------------------|------------|
| 1 | age1 | | 1 | 1 | 1227.3522 | | <.0001 |
| 2 | bmi1 | | 1 | 2 | 538.6732 | | <.0001 |
| 3 | race2 | | 1 | 3 | 104.3701 | | <.0001 |
| 4 | alcohol | | 1 | 4 | 14.0186 | | 0.0002 |
| 5 | race3 | | 1 | 5 | 13.6231 | | 0.0002 |
| 6 | age1*race2 | | 1 | 6 | 17.7187 | | <.0001 |
| 7 | bmi1*race2 | | 1 | 7 | 11.7312 | | 0.0006 |
| 8 | age1*alcohol | | 1 | 8 | 6.8738 | | 0.0087 |
| 9 | pajog | | 1 | 9 | 6.0003 | | 0.0143 |
| 10 | race2*pajog | | 1 | 10 | 4.5676 | | 0.0326 |
| 11 | gender | | 1 | 11 | 3.7650 | | 0.0523 |
| 12 | bmi1*gender | | 1 | 12 | 10.4228 | | 0.0012 |
| 13 | gender*race3 | | 1 | 13 | 9.1253 | | 0.0025 |
| 14 | hdlcholes1 | | 1 | 14 | 2.5875 | | 0.1077 |
| 15 | alcohol*pajog | | 1 | 15 | 2.4720 | | 0.1159 |

Although a definitive relationship between smoking cessation and hypertension diagnosis was not found using the biostatistical approach, this relationship could possibly be explained using the “epidemiological approach”. In the epidemiology discipline, clinical significance often plays a more important role in the analyses than statistical significance. Using the epidemiological approach, the first major difference between this method and the biostatistical method was that the continuous variables were categorized according to scientific knowledge of increasing hypertension risk within levels of each continuous variable. Secondly, analyses performed especially for continuous variables, such as the fractional polynomial analyses, were omitted from the analyses using the epidemiological approach.

The univariable analyses for the association of each variable to the dependent variable are shown in Table 13 (below). The results clearly indicate that each variable is associated with the

dependent variable ($p < .05$). Just like with the biostatistical approach, the next step was to fit a multivariable model using the variables significant at the .25 level of significance. The results, shown in Table 14 (page 38), indicate that race1 (White) and race3 (Mexican-American) may not

Table 13: Epidemiological Approach--Univariable Analyses

| Variable | Wald Chi-Square | P-value |
|-------------------|-----------------|---------|
| gender | 62.67 | <.0001 |
| White | 59.56 | <.0001 |
| Black | 54.52 | <.0001 |
| Mexican-American | 211.87 | <.0001 |
| Education | 77.55 | <.0001 |
| Age 40-59 years | 13.50 | 0.0002 |
| Age 60+ years | 1605.21 | <.0001 |
| Physical Activity | 281.10 | <.0001 |
| HDL Cholesterol | 74.78 | <.0001 |
| BMI | 492.63 | <.0001 |
| Alcohol | 211.06 | <.0001 |
| Cessation | 166.45 | <.0001 |

Table 14: Epidemiological Approach--Fitting a Preliminary Main Effects Model

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---------------------------------------|----|----------|----------------|-----------------|------------|
| Intercept | 1 | 1.0969 | 0.1738 | 39.8438 | <.0001 |
| gender Male | 1 | -0.1432 | 0.0464 | 9.5395 | 0.0020 |
| race1 White | 1 | 0.1590 | 0.1285 | 1.5325 | 0.2157 |
| race2 Not Black vs. Black | 1 | -0.6263 | 0.1303 | 23.1180 | <.0001 |
| race3 Not M-A vs. M-A | 1 | 0.1154 | 0.1328 | 0.7542 | 0.3851 |
| edu1 ≥12 years vs. <12 years | 1 | -0.0995 | 0.0492 | 4.0978 | 0.0429 |
| AGEGRP2 Not 40-49 years old | 1 | -0.9781 | 0.0569 | 295.0799 | <.0001 |
| AGEGRP3 Not 60+ years old | 1 | -1.7516 | 0.0594 | 870.9623 | <.0001 |
| pajog Phys. act. Vs. No phys. act. | 1 | -0.3105 | 0.0845 | 13.5001 | 0.0002 |
| hdl Normal HDL vs. abnormal | 1 | -0.1657 | 0.0495 | 11.1927 | 0.0008 |
| bmicat Obese vs. Not obese | 1 | 0.8598 | 0.0473 | 330.5286 | <.0001 |
| alcohol Non-drinker vs. Drinker | 1 | 0.1345 | 0.0450 | 8.9223 | 0.0028 |
| cessation Unsuccessful vs. Successful | 1 | -0.1078 | 0.0481 | 5.0330 | 0.0249 |

be significant to the model ($p > .05$). Although there is a clinical association between Black race and hypertension, there is inadequate support for the link between White race or Mexican-

American ethnicity to hypertension, therefore, the two variables are eliminated as candidates for inclusion in the final model.

Following the same procedure as the biostatistical approach, the next step in the epidemiological approach was to check for interactions. Table 15 (page 39) presents the results of fitting a model containing the main effects variables and the interactions significant at the 0.05 level of significance.

Like the biostatistical approach, the stepwise selection process was performed separately for non-interaction terms and interaction terms, a required condition using the approach suggested by Hosmer and Lemeshow. The results of applying the stepwise selection method are found in Table 16, page 40. Unlike the biostatistical method, the cessation variable was among the variables chosen for inclusion in the final model.

After taking into account the variables selected by the stepwise regression method and then considering the clinical significance of the terms, the chosen model contained the variables:

Table 15: Epidemiological Approach--Fitting a Model Containing Main Effects and Significant Interaction Terms

| Effect | DF | Wald | Pr > ChiSq |
|----------------|----|------------|------------|
| | | Chi-Square | |
| gender | 1 | 29.5235 | <.0001 |
| race2 | 1 | 109.5930 | <.0001 |
| edu1 | 1 | 0.8648 | 0.3524 |
| AGEGRP2 | 1 | 189.7646 | <.0001 |
| AGEGRP3 | 1 | 195.8407 | <.0001 |
| pajog | 1 | 15.4152 | <.0001 |
| hdl | 1 | 4.4855 | 0.0342 |
| bmicat | 1 | 3.4628 | 0.0628 |
| alcohol | 1 | 10.6399 | 0.0011 |
| cessation | 1 | 13.3887 | 0.0003 |
| gender*AGEGRP3 | 1 | 18.5174 | <.0001 |
| gender*hdl | 1 | 6.0816 | 0.0137 |
| race2*AGEGRP2 | 1 | 8.4955 | 0.0036 |
| race2*pajog | 1 | 12.1194 | 0.0005 |
| race2*bmicat | 1 | 12.5861 | 0.0004 |
| AGEGRP3*hdl | 1 | 2.2794 | 0.1311 |
| AGEGRP3*bmicat | 1 | 12.1422 | 0.0005 |

| | | | |
|------------------|---|--------|--------|
| AGEGRP3*alcohol | 1 | 1.9761 | 0.1598 |
| pajog*alcohol | 1 | 2.8378 | 0.0921 |
| bmicat*cessation | 1 | 4.4246 | 0.0354 |

agegrp3 (60+ years), bmicat, agegrp2 (40-59 years), race2 (Black), pajog, alcohol, hdl, gender, cessation, race2*pajog, race2*bmicat, and bmicat*cessation. The final model, along with the estimated coefficients, was:

$$\begin{aligned} \text{logit} = & -1.8121 \text{ agegrp3 (60+ years)} + 0.4479 \text{ bmicat (body mass index)} - .9927 \text{ agegrp2 (40-59 years)} \\ & - .7301 \text{ race2 (Black)} - .6791 \text{ pajog} + .1343 \text{ alcohol} - .1651 \text{ hdl} - .1572 \text{ gender} - .2029 \text{ cessation} \\ & + .5973 \text{ race2 * pajog} + .3136 \text{ race2 * bmicat} + .2775 \text{ bmicat * cessation} + 1.4248. \end{aligned}$$

Note that this model estimates the probability of prior hypertension diagnosis for individuals who are: not 60+ years old, obese, not 40-59 years old, non-Black, physically active, non-drinkers, at normal HDL cholesterol levels, Male, unsuccessful at smoking cessation, non-Black and physically active, non-Black and obese, and obese and unsuccessful at smoking cessation, respectively.

Table 16: Epidemiological Approach--Results of Stepwise Selection

Applying Stepwise Regression to non- interaction terms

| Step | Effect Entered | Removed | DF | Number In | Score Chi-Square | Wald Chi-Square | Pr > ChiSq |
|------|----------------|---------|----|-----------|------------------|-----------------|------------|
| 1 | AGEGRP3 | | 1 | 1 | 993.9963 | | <.0001 |
| 2 | bmicat | | 1 | 2 | 511.6015 | | <.0001 |
| 3 | AGEGRP2 | | 1 | 3 | 366.1296 | | <.0001 |
| 4 | race2 | | 1 | 4 | 140.9213 | | <.0001 |
| 5 | pajog | | 1 | 5 | 20.3869 | | <.0001 |
| 6 | alcohol | | 1 | 6 | 16.5449 | | <.0001 |
| 7 | hdl | | 1 | 7 | 7.3740 | | 0.0066 |
| 8 | gender | | 1 | 8 | 9.4954 | | 0.0021 |
| 9 | cessation | | 1 | 9 | 5.7959 | | 0.0161 |

Applying Stepwise Regression to interaction terms

| Step | Effect Entered | Removed | DF | Number In | Score Chi-Square | Wald Chi-Square | Pr > ChiSq |
|------|----------------|---------|----|-----------|------------------|-----------------|------------|
|------|----------------|---------|----|-----------|------------------|-----------------|------------|

| | | | | | |
|----|------------------|---|----|----------|--------|
| 1 | AGEGRP3 | 1 | 1 | 993.9963 | <.0001 |
| 2 | bmicat | 1 | 2 | 511.6015 | <.0001 |
| 3 | AGEGRP2 | 1 | 3 | 366.1296 | <.0001 |
| 4 | race2 | 1 | 4 | 140.9213 | <.0001 |
| 5 | pajog | 1 | 5 | 20.3869 | <.0001 |
| 6 | alcohol | 1 | 6 | 16.5449 | <.0001 |
| 7 | race2*pajog | 1 | 7 | 11.6477 | 0.0006 |
| 8 | AGEGRP3*bmicat | 1 | 8 | 10.8655 | 0.0010 |
| 9 | race2*bmicat | 1 | 9 | 10.9266 | 0.0009 |
| 10 | race2*AGEGRP2 | 1 | 10 | 8.2780 | 0.0040 |
| 11 | hdl | 1 | 11 | 6.4282 | 0.0112 |
| 12 | gender | 1 | 12 | 9.3093 | 0.0023 |
| 13 | gender*AGEGRP3 | 1 | 13 | 23.2890 | <.0001 |
| 14 | cessation | 1 | 14 | 8.1737 | 0.0043 |
| 15 | gender*hdl | 1 | 15 | 7.7017 | 0.0055 |
| 16 | bmicat*cessation | 1 | 16 | 4.5843 | 0.0323 |
| 17 | pajog*alcohol | 1 | 17 | 3.6081 | 0.0575 |

In assessing the fit and examining the model diagnostics, results were mixed (Table 17, pages 41-42, Figures 6-10, pages 42-44). Although the discrimination was good (area under the ROC curve=.744) and the Hosmer-Lemeshow goodness-of-fit p-value was non-significant (indicating a good fit), the Pearson and Deviance goodness-of-fit statistics indicated that there was not a good fit ($p<.0001$). The results of the model diagnostics reveal that the delta chi-square and delta deviance plots were indicative of good fits to the model (most of the values were less than 4).

Table 17: Epidemiological Approach--Assessing the Fit of the Model

Area Under the ROC curve

| | | | |
|--------------------|----------|-----------|--------------|
| Percent Concordant | 73.8 | Somers' D | 0.488 |
| Percent Discordant | 25.0 | Gamma | 0.494 |
| Percent Tied | 1.2 | Tau-a | 0.189 |
| Pairs | 35062544 | c | 0.744 |

Partition for the Hosmer and Lemeshow Test

| Group | Total | diagnose = 1 | | diagnose = 0 | |
|-------|-------|--------------|----------|--------------|----------|
| | | Observed | Expected | Observed | Expected |
| 1 | 1333 | 77 | 89.51 | 1256 | 1243.49 |
| 2 | 1331 | 100 | 104.63 | 1231 | 1226.37 |
| 3 | 1375 | 149 | 141.35 | 1226 | 1233.65 |
| 4 | 1339 | 224 | 218.14 | 1115 | 1120.86 |
| 5 | 1318 | 241 | 256.63 | 1077 | 1061.37 |
| 6 | 1262 | 324 | 310.55 | 938 | 951.45 |
| 7 | 1389 | 461 | 461.36 | 928 | 927.64 |
| 8 | 1326 | 525 | 496.02 | 801 | 829.98 |
| 9 | 1355 | 597 | 589.66 | 758 | 765.34 |
| 10 | 1442 | 828 | 858.16 | 614 | 583.84 |

Hosmer and Lemeshow Goodness-of-Fit Test

| Chi-Square | DF | Pr > ChiSq |
|------------|----|------------|
| 10.1853 | 8 | 0.2523 |

Classification Table

| Prob Level | Correct | | Incorrect | | | Percentages | | False POS | False NEG |
|------------|-----------|-------|-----------|-------|---------|-------------|-------------|-----------|-----------|
| | Non-Event | Event | Non-Event | Event | Correct | Sensitivity | Specificity | | |
| 0.050 | 3526 | 0 | 9944 | 0 | 26.2 | 100.0 | 0.0 | 73.8 | . |
| 0.100 | 3276 | 3240 | 6704 | 250 | 48.4 | 92.9 | 32.6 | 67.2 | 7.2 |
| 0.150 | 3153 | 3961 | 5983 | 373 | 52.8 | 89.4 | 39.8 | 65.5 | 8.6 |
| 0.200 | 2797 | 5509 | 4435 | 729 | 61.7 | 79.3 | 55.4 | 61.3 | 11.7 |
| 0.250 | 2560 | 6518 | 3426 | 966 | 67.4 | 72.6 | 65.5 | 57.2 | 12.9 |
| 0.300 | 2411 | 6843 | 3101 | 1115 | 68.7 | 68.4 | 68.8 | 56.3 | 14.0 |
| 0.350 | 1946 | 7771 | 2173 | 1580 | 72.1 | 55.2 | 78.1 | 52.8 | 16.9 |
| 0.400 | 1349 | 8686 | 1258 | 2177 | 74.5 | 38.3 | 87.3 | 48.3 | 20.0 |
| 0.450 | 1030 | 9110 | 834 | 2496 | 75.3 | 29.2 | 91.6 | 44.7 | 21.5 |
| 0.500 | 888 | 9287 | 657 | 2638 | 75.5 | 25.2 | 93.4 | 42.5 | 22.1 |
| 0.550 | 637 | 9473 | 471 | 2889 | 75.1 | 18.1 | 95.3 | 42.5 | 23.4 |
| 0.600 | 386 | 9704 | 240 | 3140 | 74.9 | 10.9 | 97.6 | 38.3 | 24.4 |

Deviance and Pearson Goodness-of-Fit Statistics

| Criterion | Value | DF | Value/DF | Pr > ChiSq |
|-----------|----------|-----|----------|------------|
| Deviance | 445.4547 | 310 | 1.4370 | <.0001 |
| Pearson | 409.7807 | 310 | 1.3219 | 0.0001 |

Number of unique profiles: 323

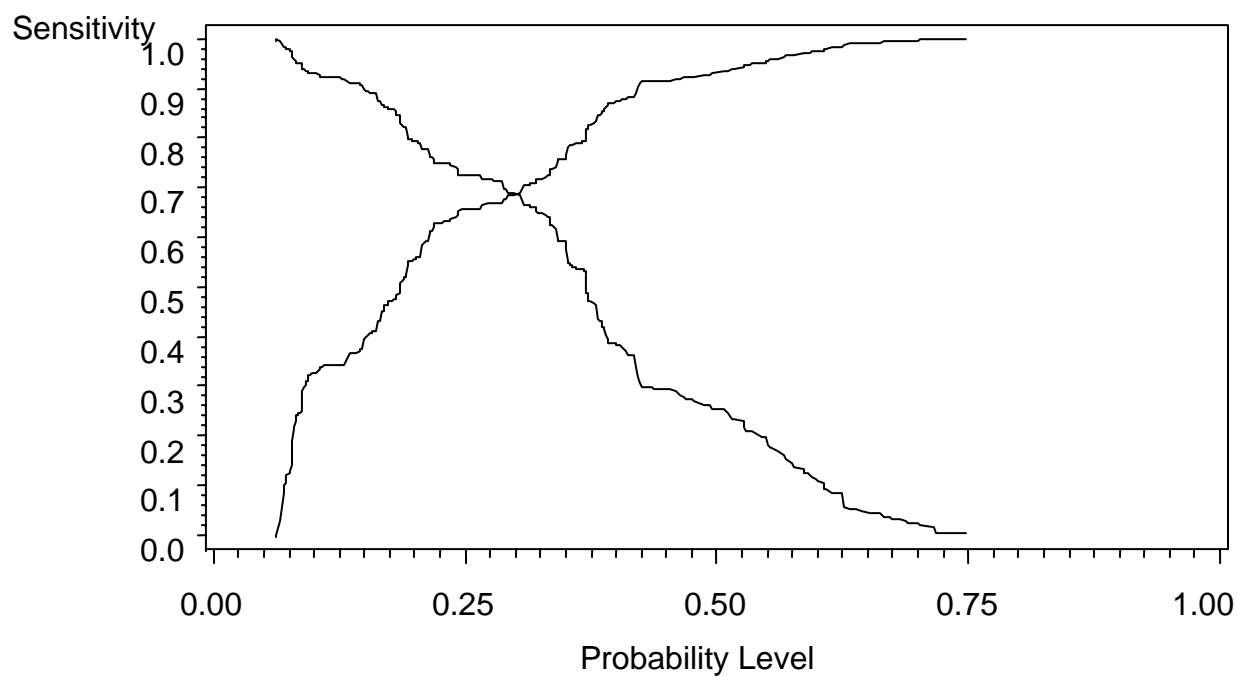


Figure 6: Epidemiological Approach--Sensitivity vs. Probability Level Plot

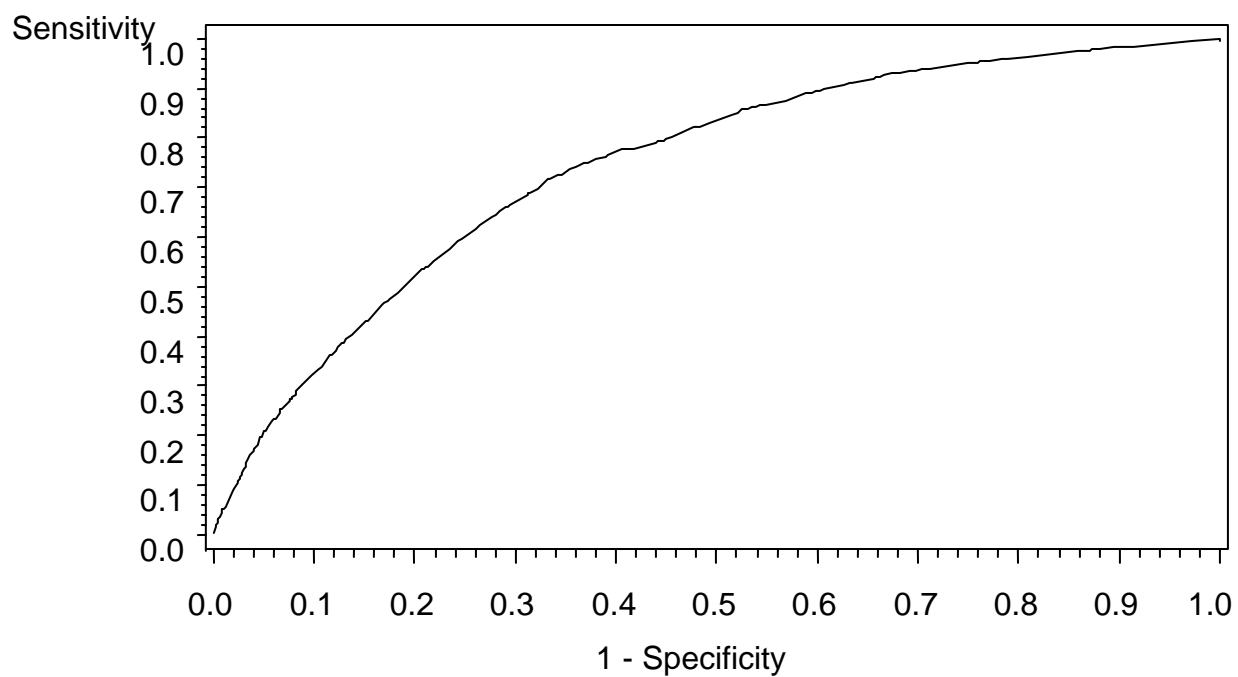


Figure 7: Epidemiological Approach—ROC Curve

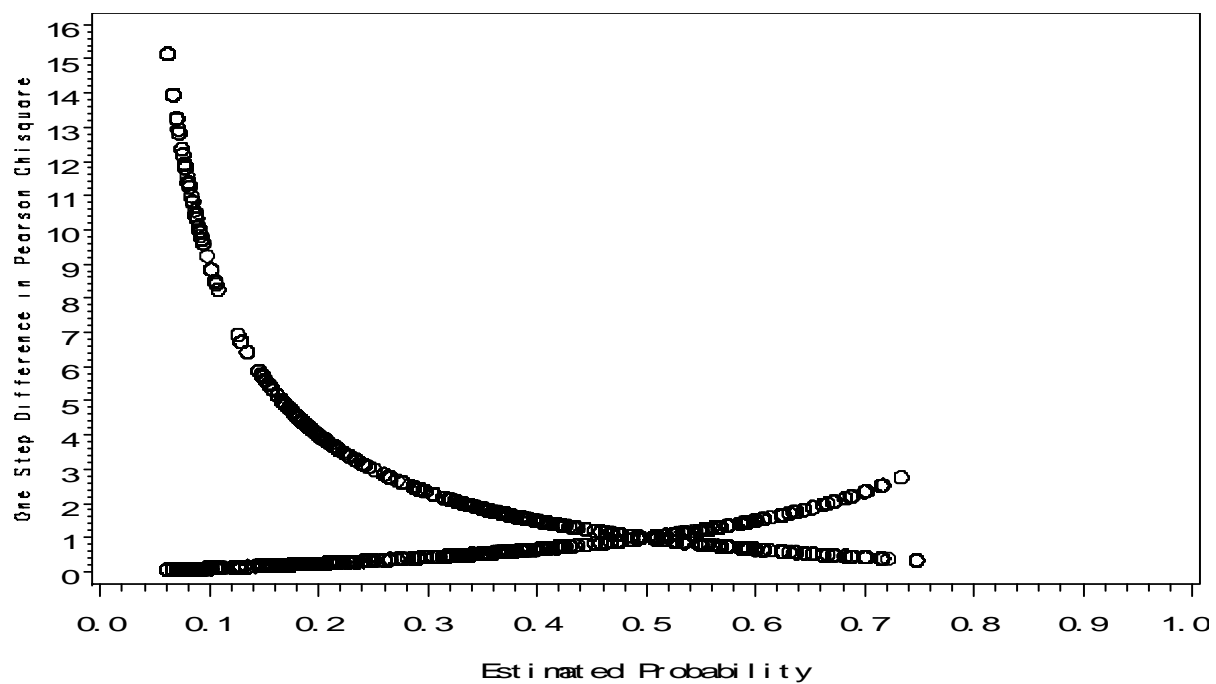


Figure 8: Epidemiological Approach--Delta Chi-Square Diagnostic Plot

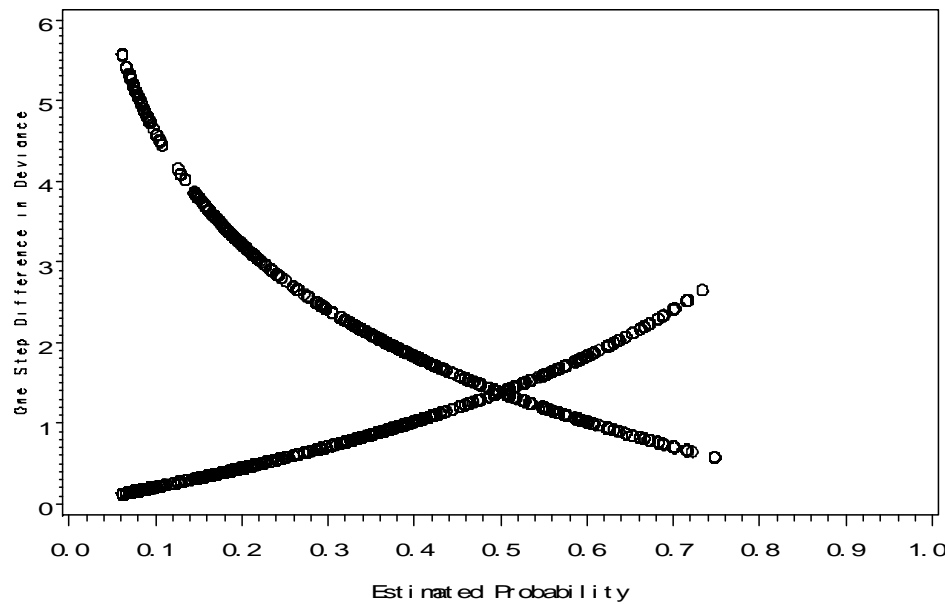


Figure 9: Epidemiological Approach--Delta Deviance Diagnostic Plot

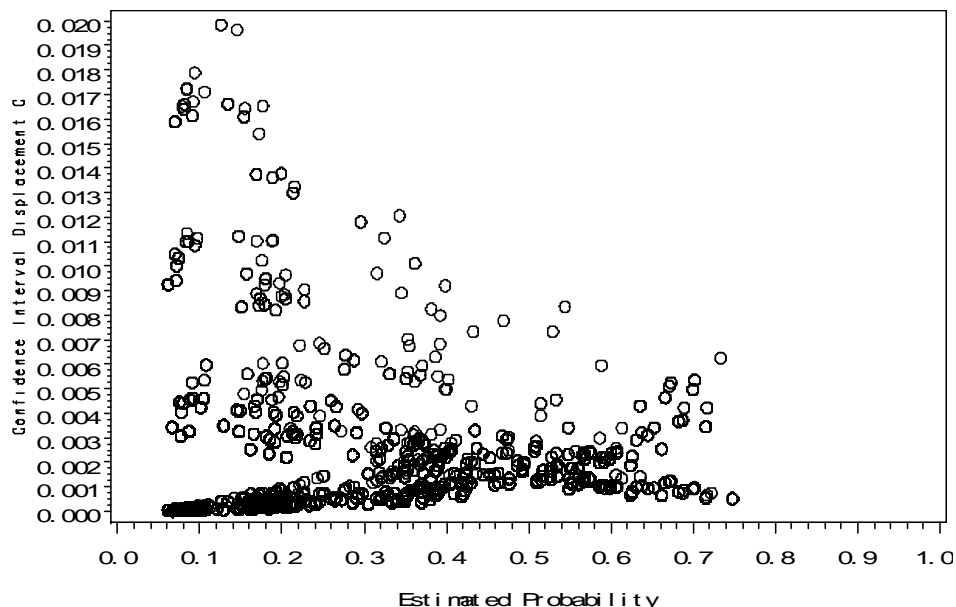


Figure 10: Epidemiological Approach--Delta Beta Hat Diagnostic Plot

Final interpretations of the model are presented in Table 18, pages 46-47. The results indicate that in comparison to those who successfully quit smoking, those who do not successfully quit are significantly less likely to have been diagnosed with hypertension, adjusting for age, bmi, race, physical activity, alcohol use, hdl cholesterol level, and gender. That is, the odds of being diagnosed with hypertension was reduced by a factor of 0.82 for unsuccessful quitters in comparison to successful quitters ($OR=.816$, $p<.001$). This means that successful quitters are at increased odds for prior hypertension diagnosis.

Other factors associated with decreased odds of having been diagnosed include: age less than 40, non-Black race, being physically active, having a normal HDL cholesterol level, and male gender. According to the model, individuals aged less than 40-59 years were 0.16 times less likely than those 40-59 years old to have been diagnosed ($OR=.163$, $p<.001$), and those aged less than 60 years old were 0.37 times less likely than those 60 years and older to have been diagnosed ($OR=0.371$, $p<.001$). Non-Blacks were nearly half as likely as Blacks ($OR=0.482$, $p<.0001$) to have

been diagnosed with hypertension. Individuals who were physically active, had normal HDL cholesterol levels, and who were male were also significantly less likely to have been diagnosed (OR=.507, .848, and .855, respectively, $p<.001$).

The estimation of the odds ratio was done separately for interaction terms. Using the method suggested by Hosmer and Lemeshow, this estimate was determined by exponentiating the difference between the odds ratio for the two levels of a covariate, evaluated at the interacting risk factor. Specifically, the odds ratio may be determined by:

$$e^{\beta_1(f_1-f_0)+\beta_3x(f_1-f_0)} = e^{[g(f_1x)-g(f_0x)]} = \frac{OR(x=1)}{OR(x=0)}.$$

For the race2*pajog interaction, the odds ratio for Black race, evaluated at the ‘no physical activity’ level was: 0.9215 and for non-Black race evaluated at the same level the odds ratio was 0.5071. The overall odds ratio of prior hypertension diagnosis for Black, physically inactive individuals was 1.817. Thus, Black, physically inactive individuals were 1.8 times more likely than non-Black, physically inactive individuals to have been diagnosed with hypertension. However, this result must be interpreted carefully because the p-value for Black race evaluated at the ‘no physical activity’ level was not significant ($p>.05$).

For the race2*bmicat interaction, the odds ratio for Black race evaluated at the ‘obese’ level was 2.145 and for the non-Black race evaluated at the same level the odds ratio was 1.5650. The overall odds ratio of prior hypertension diagnosis for Black, obese individuals was 1.3684. Thus, Black, obese individuals were 1.4 times more likely than non-Black, obese individuals to have been diagnosed with hypertension ($p<.0001$).

For the bmicat*cessation interaction, the odds ratio for unsuccessful cessation evaluated at the ‘obese’ level was 2.0655 and for successful cessation evaluated at the same level, the odds ratio

was: 1.3198. Therefore, the overall odds of hypertension diagnosis for an obese individual who is unsuccessful at cessation is 1.3 times greater than the odds for an obese individual who is successful at cessation ($p < .0001$). This suggests that when obesity is accounted for, those who successfully quit are not at greater odds of prior hypertension diagnosis, as previously indicated by the logistic regression model. Thus, the risk of obesity may play an important role in predicting successful and unsuccessful smoking cessation following diagnosis of hypertension.

Table 18: Epidemiological Approach--Final Model Interpretations
Analysis of Maximum Likelihood Estimates

| Parameter | | DF | Estimate | Standard Error | Wald Chi-Square |
|------------------|-----------------------------|----|----------|----------------|-----------------|
| Intercept | | 1 | 1.4248 | 0.1139 | 156.4345 |
| AGEGRP3 | Not 60+ years old | 1 | -1.8121 | 0.0584 | 963.8774 |
| bmicat | Obese | 1 | 0.4479 | 0.1123 | 15.9005 |
| AGEGRP2 | Not 40-49 years old | 1 | -0.9927 | 0.0567 | 306.3132 |
| race2 | Non-Black | 1 | -0.7301 | 0.0603 | 146.5915 |
| pajog | Physical activity | 1 | -0.6791 | 0.1387 | 23.9622 |
| alcohol | Non-drinker | 1 | 0.1343 | 0.0449 | 8.9650 |
| hdl | Normal Hdl cholesterol | 1 | -0.1651 | 0.0495 | 11.1203 |
| gender | Male | 1 | -0.1572 | 0.0462 | 11.5481 |
| cessation | Unsuccessful | 1 | -0.2029 | 0.0569 | 12.7285 |
| race2*pajog | Non-Black Physical activity | 1 | 0.5973 | 0.1699 | 12.3587 |
| race2*bmicat | Non-Black Obese | 1 | 0.3136 | 0.1000 | 9.8347 |
| bmicat*cessation | Obese Unsuccessful | 1 | 0.2775 | 0.0985 | 7.9291 |

Analysis of Maximum Likelihood Estimates

| Parameter | | Pr > ChiSq | Exp(Est) |
|------------------|-----------------------------|------------|----------|
| Intercept | | <.0001 | 4.157 |
| AGEGRP3 | Not 60+ years old | <.0001 | 0.163 |
| bmicat | Obese | <.0001 | 1.565 |
| AGEGRP2 | Not 40-49 years old | <.0001 | 0.371 |
| race2 | Non-Black | <.0001 | 0.482 |
| pajog | Physical activity | <.0001 | 0.507 |
| alcohol | Non-drinker | 0.0028 | 1.144 |
| hdl | Normal Hdl cholesterol | 0.0009 | 0.848 |
| gender | Male | 0.0007 | 0.855 |
| cessation | Unsuccessful | 0.0004 | 0.816 |
| race2*pajog | Non-Black Physical activity | 0.0004 | 1.817 |
| race2*bmicat | Non-Black Obese | 0.0017 | 1.368 |
| bmicat*cessation | Obese Unsuccessful | 0.0049 | 1.320 |

Wald Confidence Interval for Parameters

| Parameter | Estimate | 95% Confidence Limits | |
|---|----------|-----------------------|---------|
| Intercept | 1.4248 | 1.2015 | 1.6481 |
| AGEGRP3 Not 60+ years old | -1.8121 | -1.9265 | -1.6977 |
| bmicat Obese | 0.4479 | 0.2277 | 0.6680 |
| AGEGRP2 Not 40-49 years old | -0.9927 | -1.1039 | -0.8816 |
| race2 Non-Black | -0.7301 | -0.8483 | -0.6119 |
| pajog Physical activity | -0.6791 | -0.9511 | -0.4072 |
| alcohol Non-drinker | 0.1343 | 0.0464 | 0.2222 |
| hdl Normal Hdl cholest. | -0.1651 | -0.2621 | -0.0681 |
| gender Male | -0.1572 | -0.2478 | -0.0665 |
| cessation Unsuccessful | -0.2029 | -0.3144 | -0.0914 |
| race2*pajog Non-Black Physical activity | 0.5973 | 0.2643 | 0.9304 |
| race2*bmicat Non-Black Obese | 0.3136 | 0.1176 | 0.5096 |
| bmicat*cessation Obese Unsuccessful | 0.2775 | 0.0843 | 0.4706 |

Contrast Estimate Results

| Label | Standard Estimate | Error | Alpha | Confidence Limits | | Chi-Square | Pr>Chisq |
|--|-------------------|--------|-------|-------------------|--------|------------|----------|
| Exp(pajog=no physical activity, race2=non-Black) | 0.5071 | 0.0703 | 0.05 | 0.3863 | 0.6655 | 23.96 | <.0001 |
| Exp(pajog=no physical activity, race2=Black) | 0.9215 | 0.0947 | 0.05 | 0.7533 | 1.1271 | 0.63 | 0.4262 |
| Exp(bmicat=obese, race=Non-Black) | 1.5650 | 0.1758 | 0.05 | 1.2558 | 1.9504 | 15.90 | <.0001 |
| Exp(bmicat=obese, race=Black) | 2.1415 | 0.1788 | 0.05 | 1.8182 | 2.5223 | 83.16 | <.0001 |
| Exp(bmicat=obese, cessation=successful) | 1.5650 | 0.1758 | 0.05 | 1.2558 | 1.9504 | 15.90 | <.0001 |
| Exp(bmicat=obese, cessation=unsuccessful) | 2.0655 | 0.1761 | 0.05 | 1.7477 | 2.4412 | 72.40 | <.0001 |

Chapter V: Discussion and Conclusion

Previous studies examined the relationship between smoking and hypertension. Very few studies have considered the possibility of a link between smoking cessation and hypertension diagnosis. The specific aim of this study was to examine the potential role of undiagnosed hypertension as a barrier to smoking cessation. Understanding this association can predict future smoking behavior for smokers who are at-risk for hypertension. Such understanding may help to

isolate those risk factors associated with smoking and to modify individuals' behavior to change their odds of developing hypertension. This would be a case for primary prevention of controlling risk factors prior to the development of hypertension. By considering two methods of analyzing the data (one which emphasizes statistical significance of variables and one which emphasizes clinical significance of variables), this allows for investigating a potential relationship between smoking cessation and hypertension diagnosis in an interdisciplinary manner.

Although the present study did not identify a significant direct relationship between hypertension diagnosis and smoking cessation using the biostatistical approach, the analyses using the epidemiological approach did yield a significant relationship. That is, smokers that successfully quit were more likely to have been diagnosed with hypertension, adjusting for the presence of confounders. This supports the hypothesis of this thesis by indicating that hypertension diagnosis increases the chance of successful smoking cessation and that undiagnosed hypertension may indeed be a barrier to smoking cessation.

The epidemiological logistic regression model indicated a significant relationship between ethnicity and hypertension diagnosis. Non-Hispanic Blacks were more likely to have been diagnosed with hypertension in comparison to Whites, adjusting for the presence of confounders. This supports the results of national data which suggest that Blacks have a higher prevalence of hypertension than Whites, and that Blacks are more likely than Whites to have diagnosed hypertension.

The results of the study must be interpreted in light of limitations. One major limitation of this study is the complex sampling design and data collection methods of the NHANES data. The complexity of how the different data sets needed to be merged together resulted in many

missing observations. In addition, although sample weights, clusters, and strata were used when describing the population, they were not considered when performing the logistic regression procedures. This may or may not have lead to inaccurate conclusions.

The current smoking status reported by the individuals was another limitation of this study. Historically, people who respond to questions concerning smoking behavior have tended to underreport their smoking. If an accurate procedure were developed to measure true smoking behavior allowing the differentiation between smoker and nonsmoker, the relationship between smoking and any measurable outcome such as hypertension could be more accurately investigated.

Another limitation of this study was the way in which smoking cessation was defined. Recall that cessation was defined as having ever smoked 100 or more cigarettes and answering ‘no’ to the question ‘Do you smoke now’? Perhaps some smokers had quit a week (or even a day) prior to participation in the questionnaire. This would inaccurately measure ‘cessation’ as indicated for the purpose of this study. In many other studies, smoking cessation is defined as quitting for one year or longer and having a history of smoking 100 or more cigarettes in a lifetime. Unfortunately, defining smoking cessation in this way would have resulted in an inadequate sample size to perform the analyses. Thus, there was no way to accurately measure smokers who had quit for one year or longer. Therefore, smoking cessation as defined in this study had to be accepted based on the available data. A similar limitation of the study is that there was no way to measure when the individuals were diagnosed with hypertension. That is, it is possible for smoking cessation to have preceded the diagnosis of hypertension.

One other possible limitation of this study was the potential for recall bias in hypertension diagnosis in which individuals may have under-reported their history of diagnosis.

Also, individuals were not asked how many times they had been diagnosed. If an individual was only diagnosed once, this may not have been a true representation of their hypertension status.

It is important to note that the relationship between knowledge of a disease or condition and cessation of a risky behavior may not hold for all diseases or conditions. For example, lung cancer diagnosis may or may not lead to the decision to quit smoking. In general, individuals may be more likely to quit a modifiable risk factor (such as smoking) than a non-modifiable risk factor following diagnosis. Future studies regarding this generality are encouraged.

There are a number of recommendations for future public health efforts with respect to smoking and hypertension. In terms of smoking, previous literature suggests that most smokers attempt to quit on their own and less often use pharmacotherapy. Perhaps insurance coverage for pharmacotherapy may remove a potential financial barrier and increase the number of smokers who quit with this method of treatment. Another recommendation would be to encourage more physician involvement in smoking cessation. Provider education has been proven to increase cessation, yet in order for this to work physicians must be trained on how to incorporate smoking cessation advice into their practice.

A key recommendation for preventing hypertension would be to develop better approaches for encouraging lifestyle modifications. Educating at-risk populations about better eating habits, increasing physical activity, and quitting smoking may prove to be ineffective for populations that are unwilling to change their behavior. More emphasis should be placed on the negative consequences of hypertension, and the health beliefs of individuals in terms of their susceptibility to these consequences should be considered.

In conclusion, prevention of hypertension and increasing the number of smokers who successfully abstain from smoking are two major public health challenges. In considering

smoking as a risk factor for hypertension, life expectancy gains from helping smokers quit may be considerably greater than the benefits of treating hypertension. This thesis sought to determine whether there was a relationship between hypertension diagnosis and smoking cessation. Given the limitations of this study, the hypothesis that successful smoking cessation was associated with an increased likelihood of previous hypertension diagnosis was supported. Results also indicated that individuals who are female, Black, or over 40 years of age are more likely to have been diagnosed and are more likely to be knowledgeable about their condition. In conclusion, hypertension diagnosis may serve as a driving force in the decision to quit smoking, but given this study's limitations a clear and accurate definition of cessation is needed in future studies investigating this relationship.

References

- "Cessation and Treatment." Healthy People 2010 Goals, from
<http://www.healthypeople.gov/document/html/volume2/27Tobacco.htm>.
- . "Effectiveness of Increasing the Number of Healthcare Providers Who Advise Patients Using Tobacco to Quit." Guide to Community Preventive Services, from
www.thecommunityguide.org/tobacco.

- . "Effectiveness of Mass Media Campaigns to Reduce Initiation of Tobacco Use and Increase Cessation." Guide to Community Preventive Services, from www.thecommunityguide.org/tobacco.
 - . "Effectiveness of Telephone Counseling and Support to Help More Tobacco Users Quit." Guide to Community Preventive Services, from www.thecommunityguide.org/tobacco.
 - . "Effectiveness of Using Provider Reminders and Provider Education, With or Without Patient Education." Guide to Community Preventive Services, from www.thecommunityguide.org.
 - . "Increasing the Unit Price for Tobacco Products is Effective in Reducing Initiation of Tobacco Use and in Increasing Cessation." Guide to Community Preventive Services, from www.thecommunityguide.org/tobacco.
- Ayanian, J. Z., A. M. Zaslavsky, et al. (2003). "Undiagnosed Hypertension and Hypercholesterolemia Among Uninsured and Insured Adults in the Third National Health and Nutrition Examination Survey." American Journal of Public Health **93**(12): 2051-2054.
- Barbeau, E. M., N. Krieger, et al. (2004). "Working class matters: socioeconomic disadvantage, race/ethnicity, gender, and smoking in NHIS 2000." American Journal of Public Health **94**(2): 269-278.
- Biener, L. and M. Siegel (2000). "Tobacco Marketing and Adolescent Smoking: More Support for a Causal Inference." American Journal of Public Health **90**(3): 407-411.
- Bosworth, H. B., T. Dudley, et al. (2006). "Racial Differences in Blood Pressure Control: Potential Explanatory Factors." American Journal of Medicine **119**(1): 70-70.

Centers for Disease Control and Prevention (2005). "Cigarette Smoking Among Adults--United States, 2004." MMWR **54**(44): 1121-1124.

Centers for Disease Control and Prevention (2005). "Racial/Ethnic Disparities in Prevalence, Treatment, and Control of Hypertension--United States 1999-2002." MMWR **54**(1): 7-9.

Cronin, S. N. (1986). "Health beliefs in compliant and noncompliant hypertensive clients." Journal of Community Health Nursing **3**(2): 87-97.

Ellickson, P. L., M. Orlando, et al. (2004). "From Adolescence to Young Adulthood: Racial/Ethnic Disparities in Smoking." American Journal of Public Health **94**(2): 293-299.

England, L. J., R. J. Levine, et al. (2002). "Smoking before pregnancy and risk of gestational hypertension and preeclampsia." American Journal Of Obstetrics And Gynecology **186**(5 (Print)): 1035-1040.

Eriksen, M. P. (2000). "Best practices for comprehensive tobacco control programs: opportunities for managed care organisations." Tobacco Control **9 Suppl 1**: I11-4.

Fiore, M. C., P. A. Keller, et al. (2004). "Preventing 3 Million Premature Deaths and Helping 5 Million Smokers Quit: A National Action Plan for Tobacco Cessation" American Journal of Public Health **94**(2): 205-210.

Galvin, K. T. (1992). "A critical review of the health belief model in relation to cigarette smoking behaviour." Journal of Clinical Nursing **1**(1): 13-18.

Getliffe, K. A., R. Crouch, et al. (2000). "Hypertension awareness, detection and treatment in a university community: Results of a worksite screening." Public Health **114**(5): 361-366.

Green, L. W., M. P. Eriksen, et al. (2000). "Achieving the Implausible in the Next Decade's Tobacco Control Objectives" American Journal of Public Health **90**(3): 337-339.

- Gritz, E. R., A. V. Prokhorov, et al. (2003). "Predictors of susceptibility to smoking and ever smoking: A longitudinal study in a triethnic sample of adolescents." Nicotine & Tobacco Research **5**(4): 493-506.
- Gulliford, M. C. (2001). "Low rates of detection and treatment of hypertension among current cigarette smokers." Journal Of Human Hypertension **15**(11 (Print)): 771-773.
- Honjo, K., A. Tsutsumi, et al. (2006). "What accounts for the relationship between social class and smoking cessation? Results of a path analysis." Social Science & Medicine (1982) **62**(2 (Print)): 317-328.
- Hosmer, D., Lemeshow, S. (2000). Applied Logistic Regression, 2nd edition.
- Houston, T. K., I. C. Scarinci, et al. (2005). "Patient smoking cessation advice by health care providers: the role of ethnicity, socioeconomic status, and health." American Journal of Public Health **95**(6): 1056-1061.
- Hyland, A., Li, Q., Bauer, J., Giovino, G., Steger, C., & Cummings, K. M. (2004). Predictors of cessation in a cohort of current and former smokers followed over 13 years. Nicotine & Tobacco Research, 6(Suppl. 3), S363–S369
- Janzon, E., B. Hedblad, et al. (2004). "Changes in Blood Pressure and Body Weight Following Smoking Cessation in Women." Journal of Internal Medicine **255**: 266-72.
- Kastarinen, M., J. Tuomilehto, et al. (2002). "Smoking trends in hypertensive and normotensive Finns during 1982-1997." Journal of Human Hypertension **16**(5): 299.
- Kiefe, C. I., D. Williams, et al. (2001). "Ten-Year Changes in Smoking Among Young Adults: Are Racial Differences Explained by Socioeconomic Factors in the CARDIA Study?" American Journal of Public Health **91**(2): 213.

- King, G., A. P. Polednak, et al. (1999). "Cigarette smoking among native and foreign-born African Americans." Annals of Epidemiology **9**(4): 236-244.
- Lawlor, D. A., S. Frankel, et al. (2003). "Smoking and Ill Health: Does Lay Epidemiology Explain the Failure of Smoking Cessation Programs Among Deprived Populations?" American Journal of Public Health **93**(2): 266.
- Liszka, H. A., A. G. Mainous, III, et al. (2005). "Prehypertension and Cardiovascular Morbidity." Annals of Family Medicine **3**(4): 294-299.
- Mazzone, A., Cusa, C., Mazzucchelli, L., Vezzoli, M., Ottini, E., Ghio, S., Tossini, T., Pacifici, R., Zuccaro, P. (2001). "Cigarette Smoking and Hypertension Influence Nitric Oxide Release and Plasma Levels of Adhesion Molecules." Clinical Chemistry and Laboratory Medicine **39**(9):822-26.
- McGrady, G. A. and L. L. Pederson (2002). "Do Sex and Ethnic Differences in Smoking Initiation Mask Similarities in Cessation Behavior?" American Journal of Public Health **92**(6): 961.
- Miller, E. R., III and M. L. Jehn (2004). "New high blood pressure guidelines create new at-risk classification: changes in blood pressure classification by JNC 7." Journal of Cardiovascular Nursing **19**(6): 367-373.
- National Center for Health Statistics. National Health and Nutrition Examination Survey III: data collection forms. Hyattsville, National Center for Health Statistics: MD, USA, 1991.
- National Center for Health Statistics. Health, United States, 2005 with chart book on trends in the health of Americans. Hyattsville, National Center for

- Health Statistics: MD, USA, 2005.
- Niskanen, L., D. E. Laaksonen, et al. (2004). "Inflammation, abdominal obesity, and smoking as predictors of hypertension." Hypertension **44**(6 (Electronic)): 859-865.
- Okosun, I. S., S. Choi, et al. (2001). "Abdominal obesity defined as a larger than expected waist girth is associated with racial/ethnic differences in risk of hypertension." Journal of Human Hypertension **15**(5): 307.
- Os, I., S. Oparil, et al. (2004). "Essential hypertension in women." Blood Pressure **13**(5): 272-278.
- Panagiotakos, D. B., C. Chrysoshoou, et al. (2002). "The association between secondhand smoke and the risk of developing acute coronary syndromes, among non-smokers, under the presence of several cardiovascular risk factors: The CARDIO2000 case-control study." BMC Public Health [Electronic Resource] **2**: 9-9.
- Pardell, H., R. Tresserras, et al. (1998). "Management of the Hypertensive Patient Who Smokes." Drugs **56**(2): 177-187.
- Polednak, A. P. and G. King (1999). "Smoking Prevalence Among US Blacks: A Southern Factor?" American Journal of Health Behavior **23**(4): 280.
- Resources, G. D. o. H. (2005). "GA Division of Public Health Chemical Hazards Program." Retrieved March 10, 2006, from <http://health.state.ga.us/programs/hazards/index.asp>.
- Schmitz, J. M., R. Spiga, et al. (1999). "Smoking cessation in women with cardiac risk: A comparative study of two theoretically based therapies." Nicotine & Tobacco Research **1**(1): 87-94.

- Tonstad, S. and M. Svendsen (2005). "Premature Coronary Heart Disease, Cigarette Smoking, and the Metabolic Syndrome." American Journal of Cardiology **96**(12): 1681-1685.
- Tucker, J. S., P. L. Ellickson, et al. (2005). "Predictors of attempted quitting and cessation among young adult smokers." Preventive Medicine: An International Journal Devoted to Practice and Theory **41**(2): 554-561.
- Turner, L. R. and R. Mermelstein (2004). "Motivation and Reasons to Quit: Predictive Validity among Adolescent Smokers." American Journal of Health Behavior **28**(6): 542-550.
- United States. Public Health Service. Office of the Surgeon General. and National Center for Chronic Disease Prevention and Health Promotion (U.S.). (2004). "The health consequences of smoking: a report of the Surgeon General." from Available online, Government web site, 2004 <http://purl.access.gpo.gov/GPO/LPS49585>
- Vasan, R. S., M. G. Larson, et al. (2001). "Assessment of frequency of progression to hypertension in non-hypertensive participants in the Framingham Heart Study: a cohort study." Lancet **358**(9294): 1682.
- Wen, C. P., T. Y. Cheng, et al. (2005). "The health benefits of smoking cessation for adult smokers and for pregnant women in Taiwan." Tobacco Control **14**: i56-i61.
- Yamagishi, K., H. Iso, et al. (2003). "Smoking raises the risk of total and ischemic strokes in hypertensive men." Hypertension Research: Official Journal Of The Japanese Society Of Hypertension **26**(3 (Print)): 209-217.
- Zhang, J., M. A. Klebanoff, et al. (1999). "The puzzling association between smoking and hypertension during pregnancy." American Journal Of Obstetrics And Gynecology **181**(6 (Print)): 1407-1413.

Appendix

SAS CODES: Using the Biostatistical Approach

I. Entering / Reading the Data

A) Creating SAS datasets from the NHANES data files

```

data MYADULT;
infile 'e:\adult.dat' LRECL=3348 MISOVER;
input SEQN 1-5 DMARETHN 12 HSSEX 15 HSAGEIR 18-19 HSAGEU 20 SDPPSU6 43
SDPSTRA6 44-45 HFA8R 1256-1257 HAE2 1598 HAE5A 1610 HAR1 2281 HAR3 2285 HAR9
2300 HAT2 2396 HAZMNK1R 3337-3339 HAZMNK5R 3342-3344 WTPFQX6 52-60 WTPFEX6
61-69 WTPFHX6 70-78;
run;

data 'c:\documents and settings\owner\my documents\NHANES 1988-1994\MYADULT3';
set MYADULT;
run;

data MYEXAM2;
infile 'e:\EXAM.dat' LRECL=6235 MISOVER;
input SEQN 1-5 DMARETHN 12 HSSEX 15 HSAGEIR 16-17 HSAGEU 18 SDPPSU6 41
SDPSTRA6 42-43 BMPBMI 1524-1527 MAPE2 5110 WTPFQX6 52-60 WTPFEX6 61-69
WTPFHX6 70-78;
run;

data 'c:\documents and settings\owner\my documents\NHANES 1988-1994\MYEXAM3';
set MYEXAM2;
run;

data MYLAB2;
infile 'e:\LAB.dat' LRECL=1979 MISOVER;
input SEQN 1-5 DMARETHN 12 HSSEX 15 HSAGEIR 16-17 HSAGEU 18 SDPPSU6 41
SDPSTRA6 42-43 HDPSI 1625-1628 WTPFQX6 52-60 WTPFEX6 61-69 WTPFHX6 70-78;
run;

data 'c:\documents and settings\owner\my documents\NHANES 1988-1994\MYLAB3';
set MYLAB2;
run;

```

B) Merging the SAS 'Adult', 'Exam2', and 'Lab2' datasets

```

proc sort data="c:\documents and settings\owner\my documents\NHANES 1988-
1994\MYADULT3" out=adult;
by SEQN;
run;

proc sort data="c:\documents and settings\owner\my documents\NHANES 1988-
1994\MYEXAM3" out=exam2;
by SEQN;
run;

proc sort data="c:\documents and settings\owner\my documents\NHANES 1988-
1994\MYLAB3" out=lab2;
by SEQN;

```

```
run;
```

```
DATA "c:\documents and settings\owner\my documents\NHANES 1988-1994\MERGED3";
    MERGE exam2 lab2
        adult (IN=A);
    BY SEQN;
    IF A;
    RUN;
```

C) Coding the Variables

```
data 'f:\biostat thesis\final';
set "c:\documents and settings\owner\my documents\NHANES 1988-1994\MERGED3";
if DMARETHN =1 then race1=1; else race1=0;
if DMARETHN =2 then race2=1; else race2=0;
if DMARETHN =3 then race3=1; else race3=0;
if HSSEX=1 then gender=0;
if HSSEX=2 then gender=1;
education=HFA8R;
drop HFA8R;
if education=88 then education=.;
if education=99 then education=.;
AGE = HSAGEIR;
    IF HSAGEU = 1 THEN AGE = AGE / 12;
    IF AGE GE 20 AND AGE LE 39 THEN AGEGRP = 1;
    ELSE IF AGE GE 40 AND AGE LE 59 THEN AGEGRP = 2;
    ELSE IF AGE GE 60 THEN AGEGRP = 3;
    ELSE AGEGRP = 0;
if HAE2=2 then diagnose=0;
if HAE2=1 then diagnose=1;
if HAR1=2 then hundredcig=0;
if HAR1=1 then hundredcig=1;
if HAR3=2 then smknow=0;
if HAR3=1 then smknow=1;
if HAR9=2 then evrquit=0;
if HAR9=1 then evrquit=1;
systolic =HAZMNK1R;
if (systolic=888) then systolic= .;
diastolic =HAZMNK5R;
if (diastolic=888) then diastolic= .;
if HAT2=1 then pajog=0;
if HAT2=2 then pajog=1;
if HAE5A=2 then meds=0;
if HAE5A=1 then meds=1;
hdlcholes=HDPSI;
if (hdlcholes=8888) then hdlcholes= . ;
BMI=BMPBMI;
if (BMI=8888) then BMI= . ;
if MAPE2=2 then alcohol=0;
if MAPE2=1 then alcohol=1;
IF hundredcig="1" AND smknow="1" THEN smoker=1;
ELSE smoker=0;
IF hundredcig="1" AND smoker="0" THEN cessation=0;
ELSE cessation=1;
IF (systolic >=140) OR (diastolic >= 90) OR (meds=1) THEN hyper=1;
```

```

ELSE hyper=0;
IF hyper=1 AND diagnose=1 THEN diagstat=1;
IF hyper=1 AND diagnose=0 THEN diagstat=0;
run;

proc format;
value reth 1="Non-Hispanic White"
           2="Non-Hispanic Black"
           3="Mexican-American"
           4="Other";
value race1val 0="Non-White"
               1="White";
value race2val 0="Non-Black"
               1="Black";
value race3val 0="Non Mexican-American"
               1="Mexican-American";
value sex 0="Male"
           1="Female";
value agegroup 0="Less than 20 yrs old"
               1="20-39"
               2="40-59"
               3="60 +";
value diag 0="Not diagnosed"
            1="Diagnosed";
value cigs 0="No"
            1="Yes";
value now 0="No"
            1="Yes";
value quit 0="No"
            1="Yes";
value jog 0="Physical activity"
            1="No physical activity";
value med 0="Not taking anti-hypertensive drugs"
            1="Taking anti-hypertensive drugs";
value alc 0="Non-drinker"
            1="Drinker";
value smk 0="Non-smoker"
            1="Smoker";
value cess 0="Successful"
            1="Unsuccessful";
value hbp 0="Non-hypertensive"
            1="Hypertensive";
value diagnosed 0="Undiagnosed Hypertension"
                 1="Diagnosed Hypertension";
run;

```

C) Excluding Subjects

```

data "f:\biostat thesis\Final";
set "f:\biostat thesis\WORKINGFILE3";
if diagnose ge 0;
if cessation ge 0;
if age ge 18;
run;

```


II. Describing the Population

A) Categorical Variables: Frequencies, Test for Associations(chisq and cramersv)

i) Frequencies

```
ods output  Surveymeans.Statistics=Table1;
proc sort data='f:\biostat thesis\Final';
By SDPSTRA6 SDPPSU6;
Run;

PROC surveymeans DATA='f:\biostat thesis\Final';
    WEIGHT WTPFHx6;
    strata SDPSTRA6;
    cluster SDPPSU6;
    class DMARETHN gender agegrp diagnose pajog alcohol smoker cessation hyper
diagstat;
    VAR DMARETHN gender agegrp diagnose pajog alcohol smoker cessation hyper
diagstat;
    format DMARETHN reth. gender sex. agegrp agegroup. diagnose diag. pajog
jog. alcohol alc. smoker smk. cessation cess. hyper hbp. diagstat diagnosed.;
run;
ods trace off; ods listing close;

data final;
set Table1;
drop Mean StdErr lowerCLMean upperCLMean;
%let x=N;
if (VarName='diagstat') then Percent=&x/5375 *100; else Percent=&x/17981
*100;
Percentage=Percent;
run;
ods listing;
proc tabulate data=final order=data;
class VarName VarLevel;
var N Percentage ;
Table VarLevel, N*sum='' Percentage='% Study Population'*sum='';
Title 'Table 1: Frequency of Categorical Variables';
run;
```

ii) Testing Associations

```
proc freq data='f:\biostat thesis\final';
tables diagnose*race1 diagnose*race2 diagnose*race3 / missing chisq;
format diagnose diag. race1 race1val. race2 race2val. race3 race3val.;
run;

proc freq data='f:\biostat thesis\final';
table diagnose*gender / missing chisq;
format diagnose diag. gender sex.;
run;
proc freq data='f:\biostat thesis\final';
table diagnose*pajog / missing chisq;
```

```

format diagnose diag. pajog jog.; run;
proc freq data='f:\biostat thesis\final';
table diagnose*alcohol / missing chisq;
format diagnose diag. alcohol alc.;
run;
proc freq data='f:\biostat thesis\final';
table diagnose*cessation / missing chisq;
format diagnose diag. cessation cess.;
run;

proc freq data='f:\biostat thesis\final';
table diagnose*pajog / missing exact measures;
format diagnose diag. pajog jog.;
run;

```

B) Continuous Variables: Mean, Normality Check (skewness, kurtosis, median), Check collinearity, Check Association of Continuous Variable to Dependent Variable

i) Mean

```

proc sort data='f:\biostat thesis\final';
By SDPSTRA6 SDPPSU6;
run;
PROC surveymeans DATA='f:\biostat thesis\final';
    WEIGHT WTPFHx6;
    strata SDPSTRA6;
    cluster SDPPSU6;
    VAR Education Age Systolic Diastolic BMI Hdlcholes;
label Education='Highest Grade Completed in School';
label AGE='Age in Years';
label Systolic='Systolic Blood Pressure';
label Diastolic='Diastolic Blood Pressure';
label BMI='Body Mass Index';
label Hdlcholes='HDL Cholesterol';
run;

```

ii) Normality Check

```

proc univariate data= 'f:\biostat thesis\final';
var Education Age Systolic Diastolic Hdlcholes BMI;
label Education = 'Highest Grade Completed';
label Age= 'Age in Years';
label Systolic= 'Systolic Blood Pressure';
label Diastolic= 'Diastolic Blood Pressure';
label Hdlcholes= 'HDL cholesterol';
label BMI= 'Body Mass Index';
run;

```

iii) Check Collinearity

```

ods output Reg.MODEL1.Fit.education.ParameterEstimates=model1;
proc reg data='f:\biostat thesis\final';
model education=Age Systolic Diastolic Hdlcholes BMI/vif;
Title "Education as the dependent variable";
run;
quit;

```

```

ods output Reg.MODEL1.Fit.Age.ParameterEstimates=model2;
proc reg data='f:\biostat thesis\final';
model Age= education Systolic Diastolic Hdlcholes BMI/vif;
Title "Age as the dependent variable";
run;
quit;
ods output Reg.MODEL1.Fit.Systolic.ParameterEstimates=model3;
proc reg data='f:\biostat thesis\final';
model Systolic =Age education Diastolic Hdlcholes BMI/vif;
Title "Systolic as the dependent variable";
run;
quit;
ods output Reg.MODEL1.Fit.Diastolic.ParameterEstimates=model4;
proc reg data='f:\biostat thesis\final';
model Diastolic =Age education Systolic Hdlcholes BMI/vif;
Title "Diastolic as the dependent variable";
run;
quit;
ods output Reg.MODEL1.Fit.Hdlcholes.ParameterEstimates=model5;
proc reg data='f:\biostat thesis\final';
model Hdlcholes =Age education Systolic Diastolic BMI/vif;
Title "Hdlcholes as the dependent variable";
run;
quit;
ods output Reg.MODEL1.Fit.BMI.ParameterEstimates=model6;
proc reg data='f:\biostat thesis\final';
model BMI =Age education Systolic Diastolic Hdlcholes/vif;
Title "BMI as the dependent variable";
run;
quit;

proc print data=model1;var variable varianceinflation; run;
proc print data=model2;var variable varianceinflation; run;
proc print data=model3;var variable varianceinflation; run;
proc print data=model4;var variable varianceinflation; run;
proc print data=model5;var variable varianceinflation; run;
proc print data=model6;var variable varianceinflation; run;

```

iv) Check Association of Each Continuous Variable to the Dependent Variable

```

proc ttest data= 'f:\biostat thesis\final';
class diagnose;
var Education;
run;

proc ttest data= 'f:\biostat thesis\final';
class diagnose;
var Age;
run;

proc ttest data= 'f:\biostat thesis\final';
class diagnose;
var systolic;
run;

proc ttest data= 'f:\biostat thesis\final';

```

```
class diagnose; var diastolic; run;
proc ttest data='f:\biostat thesis\final';
class diagnose;
var Hdlcholes ;
run;
```

```
proc ttest data='f:\biostat thesis\final';
class diagnose;
var BMI ;
run;
```

III. Univariable Analyses

A) Categorical Variables

Univariable Logistic Regression for Categorical Variables

```
data ulogistic1;
set 'f:\biostat thesis\final';
proc logistic data=ulogistic1 descending;
class gender (param=ref ref='0');
model diagnose = gender / clparm=wald link=logit;
output out=sexout p=PRED lower=LPRED upper=UPRED xbeta=LOGEST
stdxbeta=LOGESTSE /alpha=0.25;
run;
```

```
data ulogistic2;
set 'f:\biostat thesis\final';
proc logistic data = ulogistic2 descending;
class race1 (param=ref ref='0');
model diagnose = race1 / clparm=wald link=logit;
output out=rethout p=PRED lower=LPRED upper=UPRED xbeta=LOGEST
stdxbeta=LOGESTSE /alpha=0.25;
run;
```

```
data ulogistic3;
set 'f:\biostat thesis\final';
proc logistic data = ulogistic3 descending;
class race2 (param=ref ref='0');
model diagnose = race2 / clparm=wald link=logit;
output out=rethout p=PRED lower=LPRED upper=UPRED xbeta=LOGEST
stdxbeta=LOGESTSE /alpha=0.25;
run;
```

```
data ulogistic4;
set 'f:\biostat thesis\final';
proc logistic data = ulogistic4 descending;
class race3 (param=ref ref='0');
model diagnose = race3 / clparm=wald link=logit;
output out=rethout p=PRED lower=LPRED upper=UPRED xbeta=LOGEST
stdxbeta=LOGESTSE /alpha=0.25; run;
```

```
data ulogistic5;
set 'f:\biostat thesis\final';
```

```

proc logistic data = ulogistic5 descending;
class alcohol (param=ref ref='0');
model diagnose = alcohol / clparm=wald link=logit;
output out=alcout p=PRED lower=LPRED upper=UPRED xbeta=LOGEST
stdxbeta=LOGESTSE /alpha=0.25;
run;

data ulogistic6;
set 'f:\biostat thesis\final';
proc logistic data=ulogistic6 descending;
class pajog (param=ref ref='0');
model diagnose = pajog / clparm=wald link=logit;
output out=hdlout p=PRED lower=LPRED upper=UPRED xbeta=LOGEST
stdxbeta=LOGESTSE /alpha=0.25;
run;

data ulogistic7;
set 'f:\biostat thesis\final';
proc logistic data = ulogistic7 descending;
class cessation (param=ref ref='0');
model diagnose = cessation / clparm=wald link=logit;
output out=cessout p=PRED lower=LPRED upper=UPRED xbeta=LOGEST
stdxbeta=LOGESTSE /alpha=0.25;
run;

```

B) Continuous Variables

i) Univariable Logistic Regression (continuous variables)

```

data contlog1;
set 'f:\biostat thesis\final';
proc logistic data=contlog1 descending;
model diagnose = age /link=logit alpha=0.25;
Title "Diagnose vs. Age";
ods select Logistic.FitStatistics;
ods select Logistic.GlobalTests;
ods select Logistic.ParameterEstimates;
run;

data contlog2;
set 'f:\biostat thesis\final';
proc logistic data=contlog2 descending;
model diagnose = education / link=logit alpha=0.25;
Title "Diagnose vs. Education";
ods select Logistic.FitStatistics;
ods select Logistic.GlobalTests;
ods select Logistic.ParameterEstimates;
run;

data contlog3;
set 'f:\biostat thesis\final';
proc logistic data=contlog3 descending;
model diagnose = systolic/ link=logit alpha=0.25;
Title "Diagnose vs. Systolic";
ods select Logistic.FitStatistics;
ods select Logistic.GlobalTests;
ods select Logistic.ParameterEstimates;

```

```

run;

data contlog4;
set 'f:\biostat thesis\final';
proc logistic data=contlog4 descending;
model diagnose = bmi/ link=logit alpha=0.25;
Title "Diagnose vs. BMI";
ods select Logistic.FitStatistics;
ods select Logistic.GlobalTests;
ods select Logistic.ParameterEstimates;
run;

data contlog5;
set 'f:\biostat thesis\final';
proc logistic data=contlog5 descending;
model diagnose = diastolic/ link=logit alpha=0.25;
Title "Diagnose vs. Diastolic";
ods select Logistic.FitStatistics;
ods select Logistic.GlobalTests;
ods select Logistic.ParameterEstimates;
run;

data contlog6;
set 'f:\biostat thesis\final';
proc logistic data=contlog6 descending;
model diagnose = hdlcholes/ link=logit alpha=0.25;
Title "Diagnose vs. HDL Cholesterol";
ods select Logistic.FitStatistics;
ods select Logistic.GlobalTests;
ods select Logistic.ParameterEstimates;
run;

```

IV. Model-Building

i) Fitting a Multivariable Model Containing Variables Significant at the .25 level

```

data multivar;
set 'f:\biostat thesis\final';
proc logistic data=multivar descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=gender race1 race2 race3 alcohol pajog cessation
age education bmi hdlcholes/alpha=.25;
run;

```

ii) Checking Linearity in the Logit: Smoothed Scatter Plots

```

data yesdiagnose;
set 'f:\biostat thesis\final';
where (diagnose=1);
run;
proc freq data=yesdiagnose;

```

```

tables age/ out=ageout;
run;
data ageplot;
set ageout;
retain y1;
if _n_ =1 then y=.29;
else y=(percent+y1);
y1=y;
run;
proc gplot data=ageplot;
plot y * age='*' / haxis=15 to 100 by 10 vaxis=0 to 100 by 10;
label y ='Percent Diagnosed';
label age='Age (Years)';
title 'Smoothed Scatter Plot of Percent Diagnosed vs. Age';
run;
quit;

proc freq data=yesdiagnose;
tables systolic/ out=sysout;
run;
data sysplot;
set sysout;
retain y1;
if _n_ <3 then y=.02;
else y=(percent+y1);
y1=y;
run;
proc gplot data=sysplot;
plot y * systolic='*' / haxis=70 to 250 by 10 vaxis=0 to 100 by 10;
label y ='Percent Diagnosed';
label systolic='Systolic Blood Pressure';
title 'Smoothed Scatter Plot of Percent Diagnosed vs. Systolic Blood
Pressure';
run;
quit;

proc freq data=yesdiagnose;
tables bmi/ out=bmiout;
run;
data bmiplot;
set bmiout;
retain y1;
if _n_ <3 then y=.02;
else y=(percent+y1);
y1=y;
run;
proc gplot data=bmiplot;
plot y * bmi='*' / haxis=10 to 90 by 5 vaxis=0 to 100 by 10;
label y ='Percent Diagnosed';
label bmi='Body Mass Index';
title 'Smoothed Scatter Plot of Percent Diagnosed vs. Body Mass Index';
run;
quit;

proc freq data=yesdiagnose;

```

```

tables diastolic/ out=diaout;
run;
proc print data=diaout;
run;
data diaplot;
set diaout;
retain y1;
if _n_ <3 then y=.02;
else y=(percent+y1);
y1=y;
run;
proc gplot data=diaplot;
plot y * diastolic='*' / haxis=10 to 150 by 10 vaxis=0 to 100 by 10;
label y = 'Percent Diagnosed';
label diastolic='Diastolic Blood Pressure';
title 'Smoothed Scatter Plot of Percent Diagnosed vs. Diastolic Blood
Pressure';
run;
quit;

proc freq data=yesdiagnose;
tables hdlcholes/ out=hdlout;
run;
proc print data=hdlout;
run;
data hdlplot;
set hdlout;
retain y1;
if _n_ <3 then y=.02;
else y=(percent+y1);
y1=y;
run;
proc gplot data=hdlplot;
plot y * hdlcholes='*' / haxis=0 to 6 by .5 vaxis=0 to 100 by 10;
label y = 'Percent Diagnosed';
label hdlcholes='HDL Cholesterol';
title 'Smoothed Scatter Plot of Percent Diagnosed vs. HDL Cholesterol';
run;
quit;

```

iiia) Fixing Non-Linearity in the Logit: Fractional Polynomial Transformation (Finding p1)

```

%macro transform(variable=);
%let x=&variable ;
data fract_poly;
set 'f:\biostat thesis\final';
y1=&x**-2;
y2=&x**-1;
y3=&x**-.5;
y4=log(&x);
y5=&x**0.5;
y6=&x;
y7=&x**2;
y8=&x**3;

```



```

ods output Genmod.ModelFit=loglik1;
data frac_poly;
set 'f:\biostat thesis\final';
y1=&x**-2;
proc genmod data=frac_poly descending;
model diagnose = y1 /dist=bin link=logit waldci;
Title "Modelling y1: &x**-2";
run;

data loglik1;
set loglik1;
if criterion='Log Likelihood';
run;

data loglik1;
set loglik1(keep=value);
run;

ods output Genmod.ModelFit=loglik2;
proc genmod data=frac_poly descending;
model diagnose = y2 /dist=bin link=logit waldci;
ods select Genmod.ModelFit;
Title "Modelling y2: &x**-1";
run;
data loglik2;
set loglik2;
if criterion='Log Likelihood';
run;

data loglik2;
set loglik2(keep=value);
run;

ods output Genmod.ModelFit=loglik3;
proc genmod data=frac_poly descending;
model diagnose = y3 /dist=bin link=logit waldci;
ods select Genmod.ModelFit;
Title "Modelling y3: &x**-0.5";
run;
data loglik3;
set loglik3;
if criterion='Log Likelihood';
run;

data loglik3;
set loglik3(keep=value);
run;

ods output Genmod.ModelFit=loglik4;
proc genmod data=frac_poly descending;
model diagnose = y4 /dist=bin link=logit waldci;
ods select Genmod.ModelFit;
Title "Modelling y4: ln &x";
run;
data loglik4;
set loglik4;

```

```

if criterion='Log Likelihood';
run;

data loglik4;
set loglik4(keep=value);
run;
ods output Genmod.ModelFit=loglik5;
proc genmod data=fract_poly descending;
model diagnose = y5 /dist=bin link=logit waldci;
ods select Genmod.ModelFit;
Title "Modelling y5: &x**0.5";
run;
data loglik5;
set loglik5;
if criterion='Log Likelihood';
run;

data loglik5;
set loglik5(keep=value);
run;

ods output Genmod.ModelFit=loglik6;
proc genmod data=fract_poly descending;
model diagnose = y6 /dist=bin link=logit waldci;
ods select Genmod.ModelFit;
Title "Modelling y6: &x";
run;
data loglik6;
set loglik6;
if criterion='Log Likelihood';
run;

data loglik6;
set loglik6(keep=value);
run;
ods output Genmod.ModelFit=loglik7;

proc genmod data=fract_poly descending;
model diagnose = y7 /dist=bin link=logit waldci;
ods select Genmod.ModelFit;
Title "Modelling y7: &x**2";
run;
data loglik7;
set loglik7;
if criterion='Log Likelihood';
run;

data loglik7;
set loglik7(keep=value);
run;

ods output Genmod.ModelFit=loglik8;
proc genmod data=fract_poly descending;
model diagnose = y8 /dist=bin link=logit waldci;
ods select Genmod.ModelFit;

```

```

Title "Modelling y8: &x**3";
run;
data loglik8;
set loglik8;
if criterion='Log Likelihood';
run;

data loglik8;
set loglik8(keep=value);
run;

data maxloglik;
set loglik1(in=in1) loglik2(in=in2) loglik3(in=in3) loglik4(in=in4)
loglik5(in=in5) loglik6(in=in6) loglik7(in=in7) loglik8(in=in8);
from1=in1; from2=in2; from3=in3; from4=in4; from5=in5; from6=in6; from7=in7;
from8=in8;run;

proc sort data=maxloglik;by descending value;run;
proc print data=maxloglik;run;

%mend transform;
run;
%transform(variable=education);run;
%transform (variable=age);run;
%transform (variable=systolic);run;
%transform (variable=bmi);run;
%transform (variable=diastolic);run;
%transform (variable=hdlcholes);run;

```

iiib) Fixing Non-Linearity in the Logit: Fractional Polynomial Transformation (Finding p2)

```

%macro transform2(variable=, p1=);
%let x=&variable ;
&p1=y1|y2|y3|y4|y5|y6|y7|y8;
data fract_poly;
set 'f:\biostat thesis\final';
y1=&x**-2;
y2=&x**-1;
y3=&x**-0.5;
y4=log(&x);
y5=&x**0.5;
y6=&x;
y7=&x**2;
y8=&x**3;
%if &p1=y1 %then %do;
ods output Genmod.ModelFit=liklil;
proc genmod data=fract_poly descending;
model diagnose = &p1 &p1*y4 / dist=bin link=logit waldci;
Title "Modelling &p1 and &p1*log(&x)";
ods select Genmod.ModelFit;run;
data liklil;
set liklil;
if criterion='Log Likelihood';
run;

```

```

data likli1;
set likli1(keep=value);
run;
%end;
%else %do;
ods output Genmod.ModelFit=likli1;
proc genmod data=fract_poly descending;
model diagnose = &p1 y1 / dist=bin link=logit waldci;
Title "Modelling &p1 and y1: &x**-2";
ods select Genmod.ModelFit;run;
data likli1;
set likli1;
if criterion='Log Likelihood';
run;

data likli1;
set likli1(keep=value);
run;
%end;
run;

%if &p1=y2 %then %do;
ods output Genmod.ModelFit=likli2;
proc genmod data=fract_poly descending;
model diagnose = &p1 &p1*y4 / dist=bin link=logit waldci;
Title "Modelling &p1 and &p1*log(&x)";
ods select Genmod.ModelFit;run;
data likli2;
set likli2;
if criterion='Log Likelihood';
run;

data likli2;
set likli2(keep=value);
run;

%end;
%else %do;
ods output Genmod.ModelFit=likli2;
proc genmod data=fract_poly descending;
model diagnose = &p1 y2 / dist=bin link=logit waldci;
Title "Modelling &p1 and y2: &x**-1";
ods select Genmod.ModelFit;run;
data likli2;
set likli2;
if criterion='Log Likelihood';
run;

data likli2;
set likli2(keep=value);
run;
%end;
run;

%if &p1=y3 %then %do;
ods output Genmod.ModelFit=likli3;

```

```

proc genmod data=fract_poly descending;
model diagnose = &p1 &p1*y4 / dist=bin link=logit waldci;
Title "Modelling &p1 and &p1*log(&x)";
ods select Genmod.ModelFit;run;
data likli3;
set likli3;
if criterion='Log Likelihood';
run;
data likli3;
set likli3(keep=value);
run;
%end;
%else %do;
ods output Genmod.ModelFit=likli3;
proc genmod data=fract_poly descending;
model diagnose = &p1 y3 / dist=bin link=logit waldci;
Title "Modelling &p1 and y3: &x**(-0.5)";
ods select Genmod.ModelFit;run;
data likli3;
set likli3;
if criterion='Log Likelihood';
run;

data likli3;
set likli3(keep=value);
run;
%end;
run;

%if &p1=y4 %then %do;
ods output Genmod.ModelFit=likli4;
proc genmod data=fract_poly descending;
model diagnose = &p1 &p1*y4 / dist=bin link=logit waldci;
Title "Modelling &p1 and &p1*log(&x)";
ods select Genmod.ModelFit;run;
data likli4;
set likli4;
if criterion='Log Likelihood';
run;

data likli4;
set likli4(keep=value);
run;

%end;
%else %do;
ods output Genmod.ModelFit=likli4;
proc genmod data=fract_poly descending;
model diagnose = &p1 y4 / dist=bin link=logit waldci;
Title "Modelling &p1 and y4: ln &x";
ods select Genmod.ModelFit;run;
data likli4;
set likli4;
if criterion='Log Likelihood';
run;

```

```

data likli4;
set likli4(keep=value);
run;

%end;
run;

%if &p1=y5 %then %do;
ods output Genmod.ModelFit=likli5;
proc genmod data=fract_poly descending;
model diagnose = &p1 &p1*y4 / dist=bin link=logit waldci;
Title "Modelling &p1 and &p1*log(&x)";
ods select Genmod.ModelFit;run;
data likli5;
set likli5;
if criterion='Log Likelihood';
run;

data likli5;
set likli5(keep=value);
run;

%end;
%else %do;
ods output Genmod.ModelFit=likli5;
proc genmod data=fract_poly descending;
model diagnose = &p1 y5 / dist=bin link=logit waldci;
Title "Modelling &p1 and y5: &x**0.5";
ods select Genmod.ModelFit;run;
data likli5;
set likli5;
if criterion='Log Likelihood';
run;

data likli5;
set likli5(keep=value);
run;
%end;
run;

%if &p1=y6 %then %do;
ods output Genmod.ModelFit=likli6;
proc genmod data=fract_poly descending;
model diagnose = &p1 &p1*y4 / dist=bin link=logit waldci;
Title "Modelling &p1 and &p1*log(&x)";
ods select Genmod.ModelFit;run;
data likli6;
set likli6;
if criterion='Log Likelihood';
run;

data likli6;
set likli6(keep=value);
run;

%end;

```

```

%else %do;
ods output Genmod.ModelFit=likli6;
proc genmod data=fract_poly descending;
model diagnose = &p1 y6 / dist=bin link=logit waldci;
Title "Modelling &p1 and y6: &x";
ods select Genmod.ModelFit;run;
data likli6;
set likli6;
if criterion='Log Likelihood';
run;
data likli6;
set likli6(keep=value);
run;
%end;
run;

%if &p1=y7 %then %do;
ods output Genmod.ModelFit=likli7;
proc genmod data=fract_poly descending;
model diagnose = &p1 &p1*y4 / dist=bin link=logit waldci;
Title "Modelling &p1 and &p1*log(&x)";
ods select Genmod.ModelFit;run;
data likli7;
set likli7;
if criterion='Log Likelihood';
run;

data likli7;
set likli7(keep=value);
run;
%end;
%else %do;
ods output Genmod.ModelFit=likli7;
proc genmod data=fract_poly descending;
model diagnose = &p1 y7 / dist=bin link=logit waldci;
Title "Modelling &p1 and y7: &x**2";
ods select Genmod.ModelFit;run;
data likli7;
set likli7;
if criterion='Log Likelihood';
run;

data likli7;
set likli7(keep=value);
run;

%end;
run;

%if &p1=y8 %then %do;
ods output Genmod.ModelFit=likli8;
proc genmod data=fract_poly descending;
model diagnose = &p1 &p1*y4 / dist=bin link=logit waldci;
Title "Modelling &p1 and &p1*log(&x)";
ods select Genmod.ModelFit;run;
data likli8;

```

```

set likli8;
if criterion='Log Likelihood';
run;

data likli8;
set likli8(keep=value);
run;

%end;
%else %do;
ods output Genmod.ModelFit=likli8;
proc genmod data=fract_poly descending;
model diagnose = &p1 y8 / dist=bin link=logit waldci;
Title "Modelling &p1 and y8: &x**8";
ods select Genmod.ModelFit;run;
data likli8;
set likli8;
if criterion='Log Likelihood';
run;

data likli8;
set likli8(keep=value);
run;
%end;
run;

data maxloglik;
set likli1(in=in1) likli2(in=in2) likli3(in=in3) likli4(in=in4)
likli5(in=in5) likli6(in=in6) likli7(in=in7) likli8(in=in8);
from1=in1; from2=in2; from3=in3; from4=in4; from5=in5; from6=in6; from7=in7;
from8=in8;run;

proc sort data=maxloglik;by descending value;run;
proc print data=maxloglik;run;

%mend transform2;
run;

%transform2(variable=education, p1=y4 )
%transform2(variable=age,p1=y3);run;
%transform2(variable=bmi,p1=y4);run;
%transform2(variable=hdlcholes,p1=y2);run;

```

iiic) Testing the Significance of p1 and p2

```

%macro test_plp2(variable=,covlist=,p1=,p2=);
data testplp2;
set 'f:\biostat thesis\final';
&p1=y1|y2|y3|y4|y5|y6|y7|y8;
&p2=y1|y2|y3|y4|y5|y6|y7|y8;
%let x=&variable ;
y1=&x**-2;
y2=&x**-1;
y3=&x**-0.5;
y4=log(&x);
y5=&x**0.5;

```



```

y6=&x;
y7=&x**2;
y8=&x**3;

proc logistic data=testplp2 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist;
Title "&x not in the model";
ods select Logistic.FitStatistics;run;

proc logistic data=testplp2 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x;
Title "&x in the model";
ods select Logistic.FitStatistics;
ods select Logistic.ParameterEstimates;run;

proc logistic data=testplp2 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &p1;
Title "Variable &x: &p1 in the model";
ods select Logistic.FitStatistics;run;

%if &p1=&p2 %then %do;
proc logistic data=testplp2 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &p1 &p1*y4;
Title "Variable &x: &p1 and log(&x)*&p1 in the model";
ods select Logistic.FitStatistics;%end;
%else %do;
proc logistic data=testplp2 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &p1 &p2;
ods select Logistic.FitStatistics;
Title "Variable &x: &p1 and &p2 in the model";%end;
run;

```

```
%mend test_plp2;
run;
```

```
%test_plp2(variable=age, covlist=gender race1 race2 race3 alcohol pajog
cessation bmi hdlcholes education,p1=y3,p2=y8);run;
```

```
%test_plp2(variable=bmi, covlist=gender race1 race2 race3 alcohol pajog
cessation age hdlcholes education,p1=y4,p2=y8);run;
```

```
%test_plp2(variable=hdlcholes, covlist=gender race1 race2 race3 alcohol pajog
cessation age bmi education,p1=y2,p2=y8);run;
```

```
%test_plp2(variable=education, covlist=gender race1 race2 race3 alcohol pajog
cessation age bmi hdlcholes,p1=y4,p2=y6);run;
```

v) Check for Interactions

```
%macro interaction1(variable=);
data int1;
set multivar;
age1=age**-0.5;
bmi1=log(bmi);
hdlcholes1=hdlcholes**-1;
education1=education;
education2=log(education);
%let x=&variable;
%let covlist=gender race1 race2 race3 alcohol pajog cessation
age1 bmi1 hdlcholes1 education1 education2;
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*race1;
ods select Logistic.ParameterEstimates;
Title "&x*race1";
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*race2;
ods select Logistic.ParameterEstimates;
Title "&x*race2";
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
```

```

        alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*race3;
ods select Logistic.ParameterEstimates;
Title "&x*race3";
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
        alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*alcohol;
ods select Logistic.ParameterEstimates;
Title "&x*alcohol";
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
        alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*pajog;
ods select Logistic.ParameterEstimates;
Title "&x*pajog";
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
        alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*cessation;
ods select Logistic.ParameterEstimates;
Title "&x*cessation";
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
        alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*age1;
ods select Logistic.ParameterEstimates;
Title "&x*age1";
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
        alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*bmil;
ods select Logistic.ParameterEstimates;
Title "&x*bmil";
run;

```

```
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*hdlcholes1;
ods select Logistic.ParameterEstimates;
Title "&x*hdlcholes1";
run;
```

```
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education1;
ods select Logistic.ParameterEstimates;
Title "&x*education1";
run;
```

```
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education2;
ods select Logistic.ParameterEstimates;
Title "&x*education2";
run;
```

```
%mend interaction1; run;
```

```
%macro interaction2(variable=);
data int1;
set multivar;
age1=age**-0.5;
bmi1=log(bmi);
hdlcholes1=hdlcholes**-1;
education1=education;
education2=log(education);
%let x=&variable;
%let covlist=gender race1 race2 race3 alcohol pajog cessation
      age1 bmi1 hdlcholes1 education1 education2;
run;
```

```
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
```

```

model diagnose=&covlist &x*alcohol;
ods select Logistic.ParameterEstimates;
Title "&x*alcohol";
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*pajog;
ods select Logistic.ParameterEstimates;
Title "&x*pajog";
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*cessation;
ods select Logistic.ParameterEstimates;
Title "&x*cessation";
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*age1;
ods select Logistic.ParameterEstimates;
Title "&x*age1";
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*bmil;
ods select Logistic.ParameterEstimates;
Title "&x*bmil";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*hdlcholes1;
ods select Logistic.ParameterEstimates;
Title "&x*hdlcholes1";
run;

```

```

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education1;
ods select Logistic.ParameterEstimates;
Title "&x*education1";
run;

```

```

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education2;
ods select Logistic.ParameterEstimates;
Title "&x*education2";
run;

```

```
%mend interaction2; run;
```

```

%macro interaction3(variable=);
data int1;
set multivar;
age1=age**-0.5;
bmil=log(bmi);
hdlcholes1=hdlcholes**-1;
education1=education;
education2=log(education);
%let x=&variable;
%let covlist=gender race1 race2 race3 alcohol pajog cessation
      age1 bmil hdlcholes1 education1 education2;
run;

```

```

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*pajog;
ods select Logistic.ParameterEstimates;
Title "&x*pajog";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;

```

```

model diagnose=&covlist &x*cessation;
ods select Logistic.ParameterEstimates;
Title "&x*cessation";
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*age1;
ods select Logistic.ParameterEstimates;
Title "&x*age1";
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*bmil;
ods select Logistic.ParameterEstimates;
Title "&x*bmil";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*hdlcholes1;
ods select Logistic.ParameterEstimates;
Title "&x*hdlcholes1";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education1;
ods select Logistic.ParameterEstimates;
Title "&x*education1";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education2;
ods select Logistic.ParameterEstimates;
Title "&x*education2";

```

```

run;

%mend interaction3; run;

%macro interaction4(variable=);
data int1;
set multivar;
age1=age**-0.5;
bmil=log(bmi);
hdlcholes1=hdlcholes**-1;
education1=education;
education2=log(education);
%let x=&variable;
%let covlist=gender race1 race2 race3 alcohol pajog cessation
              age1 bmil hdlcholes1 education1 education2;
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
          alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*cessation;
ods select Logistic.ParameterEstimates;
Title "&x*cessation";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
          alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*age1;
ods select Logistic.ParameterEstimates;
Title "&x*age1";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
          alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*bmil;
ods select Logistic.ParameterEstimates;
Title "&x*bmil";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
          alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*hdlcholes1;

```



```

ods select Logistic.ParameterEstimates;
Title "&x*hdlcholes1";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education1;
ods select Logistic.ParameterEstimates;
Title "&x*education1";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education2;
ods select Logistic.ParameterEstimates;
Title "&x*education2";
run;

%mend interaction4; run;

%macro interaction5(variable=);
data int1;
set multivar;
age1=age**-0.5;
bmi1=log(bmi);
hdlcholes1=hdlcholes**-1;
education1=education;
education2=log(education);
%let x=&variable;
%let covlist=gender race1 race2 race3 alcohol pajog cessation
age1 bmi1 hdlcholes1 education1 education2;
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*age1;
ods select Logistic.ParameterEstimates;
Title "&x*age1";
run;
proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;

```

```

format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*bmi1;
ods select Logistic.ParameterEstimates;
Title "&x*bmi1";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*hdlcholes1;
ods select Logistic.ParameterEstimates;
Title "&x*hdlcholes1";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education1;
ods select Logistic.ParameterEstimates;
Title "&x*education1";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education2;
ods select Logistic.ParameterEstimates;
Title "&x*education2";
run;

%mend interaction5; run;

%macro interaction6(variable=);
data int1;
set multivar;
age1=age**-0.5;
bmi1=log(bmi);
hdlcholes1=hdlcholes**-1;
education1=education;
education2=log(education);
%let x=&variable;
%let covlist=gender race1 race2 race3 alcohol pajog cessation
      age1 bmi1 hdlcholes1 education1 education2;
run;

```

```

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*bmi1;
ods select Logistic.ParameterEstimates;
Title "&x*bmi1";
run;

```

```

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*hdlcholes1;
ods select Logistic.ParameterEstimates;
Title "&x*hdlcholes1";
run;

```

```

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education1;
ods select Logistic.ParameterEstimates;
Title "&x*education1";
run;

```

```

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education2;
ods select Logistic.ParameterEstimates;
Title "&x*education2";
run;

```

```
%mend interaction6; run;
```

```

%macro interaction7(variable=);
data int1;
set multivar;
age1=age**-0.5;
bmi1=log(bmi);
hdlcholes1=hdlcholes**-1;
education1=education;
education2=log(education);

```

```

%let x=&variable;
%let covlist=gender race1 race2 race3 alcohol pajog cessation
              age1 bmi1 hdlcholes1 education1 education2;
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
        alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*hdlcholes1;
ods select Logistic.ParameterEstimates;
Title "&x*hdlcholes1";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
        alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education1;
ods select Logistic.ParameterEstimates;
Title "&x*education1";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
        alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education2;
ods select Logistic.ParameterEstimates;
Title "&x*education2";
run;

%mend interaction7; run;

%macro interaction8(variable=);
data int1;
set multivar;
age1=age**-0.5;
bmi1=log(bmi);
hdlcholes1=hdlcholes**-1;
education1=education;
education2=log(education);
%let x=&variable;
%let covlist=gender race1 race2 race3 alcohol pajog cessation
              age1 bmi1 hdlcholes1 education1 education2;
run;

proc logistic data=int1 descending;

```

```

class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education1;
ods select Logistic.ParameterEstimates;
Title "&x*education1";
run;

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=&covlist &x*education2;
ods select Logistic.ParameterEstimates;
Title "&x*education2";
run;
%mend interaction8; run;

%interaction1(variable=gender);run;

%interaction2(variable=race1);run;

%interaction2(variable=race2);run;

%interaction2(variable=race3);run;

%interaction3(variable=alcohol);run;

%interaction4(variable=pajog);run;

%interaction5(variable=cessation);run;

%interaction6(variable=age1);run;

%interaction7(variable=bmi1);run;

%interaction8(variable=hdlcholes1);run;

```

V) Model-Selection

a) Stepwise Selection Method

```

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
      alcohol alc. pajog jog. cessation cess.;
model diagnose=gender race1 race2 race3 alcohol pajog cessation
      age bmi hdlcholes education education/selection=stepwise
      slentry=.15 slstay=.20 details;
Title "Applying Stepwise Regression to non-interaction terms";
run;

```

```

proc logistic data=int1 descending;
class gender (ref=first) race1(ref=first)
race2(ref=first) race3(ref=first) alcohol(ref=first)
pajog(ref=first) cessation(ref=first) /param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
alcohol alc. pajog jog. cessation cess.;
model diagnose=gender race1 race2 race3 alcohol pajog cessation
age1 bmil hdlcholes1 education1 education2 bmil*race2
education1*race2 age1*race3 alcohol*pajog age1*alcohol age1*cessation
bmil*hdlcholes1 education2*race1 race2*pajog age1*race2 gender*race2
gender*race3 bmil*gender education1*gender race1*pajog age1*race1 bmil*race1
/selection=stepwise
slentry=.15 slstay=.20 details;
Title "Applying Stepwise Regression to interaction terms";
run;

```

SAS CODES: Using the Epidemiological Approach

I. Creating the dataset

A) Coding the Variables

```

data 'f:\biostat thesis\final2';
set 'f:\biostat thesis\final';
if DMARETHN =1 then race1=1; else race1=0;
if DMARETHN =2 then race2=1; else race2=0;
if DMARETHN =3 then race3=1; else race3=0;
if HSSEX=1 then gender=0;
if HSSEX=2 then gender=1;
if education=88 then education=.;
if education=99 then education=.;
if education<=12 thenedul=1; else edul=0;
AGE = HSAGEIR;
IF HSAGEU = 1 THEN AGE = AGE / 12;
IF AGE GE 40 AND AGE LE 59 THEN AGEGRP2 = 1; ELSE AGEGRP2 = 0;
IF AGE GE 60 THEN AGEGRP3 = 1; ELSE AGEGRP3 = 0;
if HAE2=2 then diagnose=0;
if HAE2=1 then diagnose=1;
if HAR1=2 then hundredcig=0;
if HAR1=1 then hundredcig=1;
if HAR3=2 then smknow=0;
if HAR3=1 then smknow=1;
if HAR9=2 then evrquit=0;
if HAR9=1 then evrquit=1;
systolic =HAZMKN1R;
if (systolic=888) then systolic= .;
if systolic>=140 then sys=1; else sys=0;
diastolic =HAZMKN5R;
if (diastolic=888) then diastolic= .;

```

```

if diastolic>=90 then dia=1; else dia=0;
if HAT2=1 then pajog=0;
if HAT2=2 then pajog=1;
if HAE5A=2 then meds=0;
if HAE5A=1 then meds=1;
hdlcholes=HDPSI;
if (hdlcholes=8888) then hdlcholes=. ;
if hdlcholes>=1 then hdl=0; else hdl=1;
BMI=BMPBMI;
if (BMI=8888) then BMI=. ;
if bmi>=30 then bmicat=1; else bmicat=0;
if MAPE2=2 then alcohol=0;
if MAPE2=1 then alcohol=1;
IF hundredcig="1" AND smknow="1" THEN smoker=1;
ELSE smoker=0;
IF hundredcig="1" AND smoker="0" THEN cessation=0;
ELSE cessation=1;
IF (systolic >=140) OR (diastolic >= 90) OR (meds=1) THEN hyper=1;
ELSE hyper=0;
IF hyper=1 AND diagnose=1 THEN diagstat=1;
IF hyper=1 AND diagnose=0 THEN diagstat=0; run;

```

```

proc format;
value reth 1="Non-Hispanic White"
           2="Non-Hispanic Black"
           3="Mexican-American"
           4="Other";
value race1val 0="Non-White"
               1="White";
value race2val 0="Non-Black"
               1="Black";
value race3val 0="Non Mexican-American"
               1="Mexican-American";
Value edu1val 0="Not high-school level education"
              1="High school level education";
Value edu2val 0="Not undergraduate level education"
              1="Undergraduate level education";
value sex 0="Male"
          1="Female";
value age1group 0="Not 20-39 years old"
               1="20-39 years old";
Value age2group 0="Not 40-49 years old"
               1="40-59 years old";
Value age3group 0="Not 60+ years old"
               1="60 + years old";
value diag 0="Not diagnosed"
           1="Diagnosed";
Value sys 0="Non-Hypertensive systolic blood pressure"
          1="Hypertensive systolic blood pressure";
Value dia 0="Non-Hypertensive diastolic blood pressure"
          1="Hypertensive diastolic blood pressure";
Value hdl 0="Normal Hdl cholesterol"
          1="Abnormal Hdl cholesterol";
Value bmicat 0="Not obese"
             1="Obese";
value cigs 0="No"

```

```

        1="Yes";
value now 0="No"
        1="Yes";
value quit 0="No"
        1="Yes";
value jog 0="Physical activity"
        1="No physical activity";
value med 0="Not taking anti-hypertensive drugs"
        1="Taking anti-hypertensive drugs";
value alc 0="Non-drinker"
        1="Drinker";
value smk 0="Non-smoker"
        1="Smoker";
value cess 0="Successful"
        1="Unsuccessful";
value hbp 0="Non-hypertensive"
        1="Hypertensive";
value diagnosed 0="Undiagnosed Hypertension"
        1="Diagnosed Hypertension";
run;

```

B) Excluding Subjects

```

data "f:\biostat thesis\final2";
set "f:\biostat thesis\final2";
if diagnose ge 0;
if cessation ge 0;
if age ge 18;
run;

```

II. Univariable Analyses

A) Categorical Variables

Univariable Logistic Regression for Categorical Variables

```

%macro univariable(variable=);
%let x=&variable;
data ulogistic;
set 'f:\biostat thesis\final2';
proc logistic data=ulogistic descending;
class &variable (param=ref ref='0');
model diagnose = &variable / clparm=wald link=logit;
output out=sexout p=PRED lower=LPRED upper=UPRED xbeta=LOGEST
stdxbeta=LOGESTSE /alpha=0.25;
ods select Logistic.FitStatistics;
ods select Logistic.GlobalTests;
Title "Univariable logistic regression of &x"
run;
%mend univariable;
Run;

%univariable(variable=gender);
%univariable(variable=race1);

```



```
%univariable(variable=race2);
%univariable(variable=race3);
%univariable(variable=edul);
%univariable(variable=agegrp1);
%univariable(variable=agegrp2);
%univariable(variable=agegrp3);
%univariable(variable=sys);
%univariable(variable=dia);
%univariable(variable=pajog);
%univariable(variable=hdl);
%univariable(variable=bmicat);
%univariable(variable=alcohol);
%univariable(variable=cessation);
```

III. Model-Building

i) Fitting a Multivariable Model Containing Variables Significant at the .25 level

```
data multivar;
set 'f:\biostat thesis\final2';
proc logistic data=multivar descending;
model diagnose=gender race1 race2 race3 edul agegrp2 agegrp3
pajog hdl bmocat alcohol cessation /alpha=.25;
run;
```

IV) Check for Interactions

```
%macro interaction1(variable=);
data int1;
set 'f:\biostat thesis\final2';
%let x=&variable;
%let covlist= gender race2 agegrp2 agegrp3
pajog hdl bmocat edul alcohol cessation;
run;
proc logistic data=int1 descending;
class gender(ref=first) race2(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race2 race2val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*race2;
ods select Logistic.ParameterEstimates;
Title "&x*race2";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*agegrp2;
```

```

ods select Logistic.ParameterEstimates;
Title "&x*agegrp2";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*agegrp3;
ods select Logistic.ParameterEstimates;
Title "&x*agegrp3";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*pajog;
ods select Logistic.ParameterEstimates;
Title "&x*pajog";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*hdl;
ods select Logistic.ParameterEstimates;
Title "&x*hdl";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*bmicat;
ods select Logistic.ParameterEstimates;
Title "&x*bmicat";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;

```

```

format gender sex. race1 race1val. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*alc;
ods select Logistic.ParameterEstimates;
Title "&x*alc";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*cessation;
ods select Logistic.ParameterEstimates;
Title "&x*cessation";
run;
%mend interaction1; run;

%macro interaction2(variable=);
data int1;
set 'f:\biostat thesis\final2';
%let x=&variable;
%let covlist= gender race2 agegrp2 agegrp3
pajog hdl bmocat edul alcohol cessation;
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*agegrp2;
ods select Logistic.ParameterEstimates;
Title "&x*agegrp2";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 race1val. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*agegrp3;
ods select Logistic.ParameterEstimates;
Title "&x*agegrp3";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)

```

```

cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*pajog;
ods select Logistic.ParameterEstimates;
Title "&x*pajog";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*hdl;
ods select Logistic.ParameterEstimates;
Title "&x*hdl";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*bmocat;
ods select Logistic.ParameterEstimates;
Title "&x*bmocat";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*alc;
ods select Logistic.ParameterEstimates;
Title "&x*alc";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*cessation;
ods select Logistic.ParameterEstimates;
Title "&x*cessation";
run;

```

```

%mend interaction2; run;

%macro interaction3(variable=);
data int1;
set 'f:\biostat thesis\final2';
%let x=&variable;
%let covlist= gender race2 agegrp2 agegrp3
pajog hdl bmicat edul alcohol cessation;
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*pajog;
ods select Logistic.ParameterEstimates;
Title "&x*pajog";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*hdl;
ods select Logistic.ParameterEstimates;
Title "&x*hdl";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*bmicat;
ods select Logistic.ParameterEstimates;
Title "&x*bmicat";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*alcohol;
ods select Logistic.ParameterEstimates;
Title "&x*alcohol";

```

```

run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*cessation;
ods select Logistic.ParameterEstimates;
Title "&x*cessation";
run;
%mend interaction3; run;

```

```

%macro interaction4(variable=);
data int1;
set 'f:\biostat thesis\final2';
%let x=&variable;
%let covlist= gender race2 agegrp2 agegrp3
pajog hdl bmicat edul alcohol cessation;
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*hdl;
ods select Logistic.ParameterEstimates;
Title "&x*hdl";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*bmicat;
ods select Logistic.ParameterEstimates;
Title "&x*bmicat";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*alc;

```

```

ods select Logistic.ParameterEstimates;
Title "&x*alc";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*cessation;
ods select Logistic.ParameterEstimates;
Title "&x*cessation";
run;
%mend interaction4; run;

%macro interaction5(variable=);
data int1;
set 'f:\biostat thesis\final2';
%let x=&variable;
%let covlist= gender race2 agegrp2 agegrp3
pajog hdl bmicat edul alcohol cessation;
run;

proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*bmicat;
ods select Logistic.ParameterEstimates;
Title "&x*bmicat";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*alc;
ods select Logistic.ParameterEstimates;
Title "&x*alc";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.

```

```

bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*cessation;
ods select Logistic.ParameterEstimates;
Title "&x*cessation";
run;
%mend interaction5; run;

%macro interaction6(variable=);
data int1;
set 'f:\biostat thesis\final2';
%let x=&variable;
%let covlist= gender race2 agegrp2 agegrp3
pajog hdl bmocat edul alcohol cessation;
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*alc;
ods select Logistic.ParameterEstimates;
Title "&x*alc";
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*cessation;
ods select Logistic.ParameterEstimates;
Title "&x*cessation";
run;
%mend interaction6; run;

%macro interaction7(variable=);
data int1;
set 'f:\biostat thesis\final2';
%let x=&variable;
%let covlist= gender race2 agegrp2 agegrp3
pajog hdl bmocat edul alcohol cessation;
run;
proc logistic data=int1 descending;
class gender(ref=first) race1(ref=first) race2(ref=first) race3(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race1 racelval. race2 race2val. race3 race3val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=&covlist &x*cessation;

```



```
ods select Logistic.ParameterEstimates;
Title "&x*cessation";
run;
%mend interaction7; run;
```

```
%interaction1(variable=gender);run;
```

```
%interaction2(variable=race2);run;
```

```
%interaction3(variable=agegrp2);run;
```

```
%interaction3(variable=agegrp3);run;
```

```
%interaction4(variable=pajog);run;
```

```
%interaction5(variable=hdl);run;
```

```
%interaction6(variable=bmicat);run;
```

```
%interaction7(variable=alcohol);run;
```

V) Fitting a Model Containing the Main Effects + Interactions Significant at .05 level (to see which interaction terms are important to the model)

```
proc logistic data=multivar descending;
class gender(ref=first) race2(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race2 race2val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=gender race2 edul agegrp2 agegrp3
pajog hdl bmocat alcohol cessation
gender*agegrp3 gender*hdl race2*agegrp2 race2*pajog race2*bmicat agegrp3*hdl
agegrp3*bmicat agegrp3*alcohol pajog*alcohol bmocat*cessation/alpha=.25;
run;
```

VI) Model-Selection

a) Stepwise Selection Method

```
proc logistic data=multivar descending;
class gender(ref=first) race2(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmocat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race2 race2val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmocat bmocat. alcohol alc. cessation cess.;
model diagnose=gender race2 edul agegrp2 agegrp3
```

```
pajog hdl bmicat alcohol cessation/selection=stepwise slentry=.15 slstay=.20
details;
Title "Applying Stepwise Regression to non-interaction terms";
run;
```

```
proc logistic data=multivar descending;
class gender(ref=first) race2(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race2 race2val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=gender race2 edul agegrp2 agegrp3
pajog hdl bmicat alcohol cessation
gender*agegrp3 gender*hdl race2*agegrp2 race2*pajog race2*bmicat agegrp3*hdl
agegrp3*bmicat agegrp3*alcohol pajog*alcohol
bmicat*cessation/selection=stepwise slentry=.15 slstay=.20 details;
Title "Applying Stepwise Regression to interaction terms";
run;
```

VII) Assessing the Fit of the Model

a) Area under ROC curve, HL Goodness of Fit, Classification Tables, ROC plot

```
proc logistic data=multivar descending;
class gender(ref=first) race2(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race2 race2val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=agegrp3 bmicat agegrp2 race2 pajog alcohol hdl gender
cessation
race2*pajog race2*bmicat
bmicat*cessation/details lackfit
clparm=wald link=logit ctable outroc=finalmodel pprob = (.05 to .6 by .05);
Title "Assessing the Fit of the Model";
run;

data finalmodel2;
set finalmodel;
spec = 1-_lmspec_;
run;
symbol1 i=join v=none ;
proc gplot data=finalmodel2;
plot _sensit*_PROB_=1 spec*_PROB_=1 / overlay haxis=0 to 1 by .25 vaxis=0
to 1 by .1 ;
run;
quit;

symbol1 i=join v=none ;
proc gplot data=finalmodel;
title 'ROC Curve';
```

```

    plot _sensit*_lmspec_=1 / vaxis=0 to 1 by .1 ;
run;
quit;
title;

```

b) Other Measures of Goodness of Fit

i) Pearson Chi-Square and Deviance Summary Statistics

```

data premodel2;
set multivar;
race2jog= race2*pajog;
race2bmi= race2*bmicat ;
bmices= bmicat*cessation;
;
run;
proc sort data=premodel2 out=premodsort nodupkey;
    by agegrp3 bmicat agegrp2 race2 pajog alcohol hdl gender cessation
    race2jog    race2bmi
    bmices;
run;
proc sort data=premodel2 out=premodsorta;
    by agegrp3 bmicat agegrp2 race2 pajog alcohol hdl gender cessation
    race2jog    race2bmi
    bmices;

run;
data premod53;
    set premodsort;
    covpat=_n_;
run;
data premod54;
    merge premodsorta premod53;
    by agegrp3 bmicat agegrp2 race2 pajog alcohol hdl gender cessation
    race2jog    race2bmi
    bmices;
run;
proc logistic data=premod54 desc;
class gender(ref=first) race2(ref=first)
    edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
    pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
    cessation(ref=first)/param=ref;
    model diagnose = agegrp3 bmicat agegrp2 race2 pajog alcohol hdl gender
    cessation
    race2jog    race2bmi
    bmices
/ aggregate scale = 1;
    output out=premod55 p=estprob DIFCHISQ=deltachi DIFDEV=deltad c=deltabeta
h=lev;
run;

```

VIII) Model Diagnostics

a) Delta Chi-Square vs. Estimated Probability plot

```

proc logistic data=premodel2 desc noprint;
  model diagnose = agegrp3 bmicat agegrp2 race2 pajog alcohol hdl gender
  cessation
  race2jog race2bmi bmicess
  / aggregate scale = 1;
  output out=premod52 p=estprob DIFCHISQ=deltachi DIFDEV=deltad c=deltabeta
  h=lev;
run;
symbol1 i=none v=circle ;
axis1 order=(0 to 1 by .1) ;
axis2 label=(angle=90 color=black height=0.75);
proc gplot data=premod52;
  plot deltach*estprob=1 / haxis=axis1 vaxis=axis2 ;
run;
quit;

```

b) Delta Deviance vs. Estimated Probability plot

```

symbol1 i=none v=circle ;
axis1 order=(0 to 1 by .1) ;
axis2 label=(angle=90 color=black height=0.75);
proc gplot data=premod52;
  plot deltad*estprob=1 / haxis=axis1 vaxis=axis2;
run; quit;

```

c) Delta Beta hat vs. Estimated Probability plot

```

symbol1 i=none v=circle ;
axis1 order=(0 to 1 by .1) ;
axis2 label=(angle=90 color=black height=0.75);
proc gplot data=premod52;
  plot deltabeta*estprob=1 / haxis=axis1 vaxis=axis2;
run;
quit;

```

IX) Final Model Interpretations**a) Estimated Coefficients, Standard Errors, 95% Confidence Intervals and Odds Ratios**

```

proc logistic data=multivar descending;
  class gender(ref=first) race2(ref=first)
  edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
  pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
  cessation(ref=first)/param=ref;
  format gender sex. race2 race2val.
  edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
  bmicat bmicat. alcohol alc. cessation cess.;
  model diagnose=agegrp3 bmicat agegrp2 race2 pajog alcohol hdl gender
  cessation
  race2*pajog race2*bmicat
  bmicat*cessation/expb clparm=wald;
run;

```

b) Odds Ratios for Interaction Terms

```

proc genmod data=premodel2 descending;
class gender(ref=first) race2(ref=first)
edul(ref=first) agegrp2(ref=first) agegrp3 (ref=first)
pajog(ref=first) hdl(ref=first) bmicat(ref=first) alcohol(ref=first)
cessation(ref=first)/param=ref;
format gender sex. race2 race2val.
edul edulval. agegrp2 age2group. agegrp3 age3group. pajog jog. hdl hdl.
bmicat bmicat. alcohol alc. cessation cess.;
model diagnose=agegrp3 bmicat agegrp2 race2 pajog alcohol hdl gender
cessation
race2*pajog race2*bmicat
bmicat*cessation
/ dist=bin link=logit waldci;
estimate 'pajog=no physical activity, race2=non-Black' pajog 1 race2*pajog
0/exp;
estimate 'pajog=no physical activity, race2=Black' pajog 1 race2*pajog
1/exp;
estimate 'bmicat=obese, race=Non-Black' bmicat 1 race2*bmicat 0/exp;
estimate 'bmicat=obese, race=Black' bmicat 1 race2*bmicat 1/exp;
estimate 'bmicat=obese, cessation=successful' bmicat 1 bmicat*cessation
0/exp;
estimate 'bmicat=obese, cessation=unsuccessful' bmicat 1 bmicat*cessation
1/exp;
run;

```